

Deciding where the money goes

The US Congress is locked in its annual struggle over the budget. Although the outcome is not yet clear, there are good reasons why researchers should take a closer interest.

THE annual muddle in the United States over who gets what piece of the budget pie is reaching its final few frantic weeks. That the 1988 fiscal year has started without a single appropriation bill signed into law has caused virtually no stir in Washington. Indeed, many seem unaware that the 1 October starting date has come and gone. The Congress, some nine months after it received the 1988 budget blueprint from the White House, has managed only to give itself an additional 45 days to decide what to do about it. The perennial difficulty of Congress failing to decide in good time how the government's coffers should be emptied is one thing. The uncertainty this causes for agency planners in Washington is another, almost an invitation to maladministration, which this year is exacerbated by the prospect that the revised Gramm–Rudman deficit reduction process may be invoked. Nobody seems to know what that will do to budgets, except reduce them.

What makes the US budget process so complex? Mostly it is the number of players involved. The agencies within the Reagan Administration devise budget plans that are reviewed by the White House Office of Management and Budget. After an intense but generally well hidden internal debate, the president delivers a budget blueprint to Congress. Once this arrives on Capitol Hill, the budget is fair game. Congressmen with political, economic and social axes to grind are let loose on the government spending plans, prodded by lobbyists from every imaginable interest group. The Washington telephone directory is a reminder that bituminous contractors as well as biotechnology companies feel the need for associations near the seat of government. A plethora of hearings held by a dizzying array of committees and subcommittees is designed to elicit the opinion of the entire spectrum of the American public on how their tax dollars should be spent. Outside the hearing rooms, congressmen hear informally and often more persuasively about the needs of their constituents and other interested parties.

It is little wonder, given the diversity of interests that must be reconciled, that the process is tortuous. In past years, Congress has found it impossible to resolve its budgetary dilemmas until the absolute last moment, with the government actually going through the absurd process of closing the buildings and sending its employees home because Congress has not provided more spending money in time. This has prompted a search for some mechanism that would cut through this budgetary logjam. One, so far not adopted, would amend the constitution to require Congress and the president to agree on a balanced budget. The expectation that constitutional fiat could accomplish what legislation has failed to do is touching evidence of the power people attach to that document, but supporters of the amendment have failed to anticipate the shattering crisis there would be if even an amended constitution failed to produce a spending plan on schedule. As yet, the idea is a long way from reality.

Another attack on business as usual has come from the Gramm–Rudman deficit reduction act. This law sets deadlines for Congress and the president in proceeding through the budget maze. If deadlines are missed, budgets are cut automatically and across the board — although even here there are exceptions for some programmes with sacrosanct spending

authority. The automatic cuts prescribed by the Gramm–Rudman act represent the ultimate abdication of responsibility by the government, an admission that the rational approach to setting a budget has failed, leaving the governing process to unsophisticated computer programs that need only to do simple multiplication. It is easy to blame Congress alone for having arrived at this unfortunate state, but responsibility also falls on the White House for refusing to compromise on spending, and threatening to veto a budget with which it is not happy.

The temptation for researchers sitting comfortably in university laboratories who read about these goings on is either to shake their heads sadly or to congratulate themselves for not having to be involved in such a mess. It is not inconsistent to do both. But this temptation must be resisted, for the shenanigans in Washington are not just part of some droll revue. The hearings, meetings and political dealings will ultimately decide the fate of all those comfortable laboratories and universities. The steps required to arrive at a budget may seem ludicrous when viewed from the outside, as indeed they do for many on the inside, but that is how the game is played. To ignore the process, to wait for the dust to clear, is a certain formula for being left out when the resources grow smaller and the number of people seeking a share grows larger.

Congressmen have already shown a tendency to grow impatient with experts appearing before them who act aloof or uninterested; that impatience could easily turn into hostility. This is not to say that part of becoming a scientist should include practising obeisances and learning to kowtow. Rather, researchers and others involved in spending the government's money must inject themselves into the decision making process with vigour and enthusiasm. This can mean letter writing, attending public forums, visiting congressmen or other well-trodden routes for participating in the process of government.

In addition to making a case for strong support of science to elected officials, there is also a need to reach the electors. Scientists have an obligation, as they are spending the public's money, to explain just what they are doing with it. At times, explanation will be difficult; there is a lot of new vocabulary that must be presented to a lay audience before the concept of a restriction fragment length polymorphism can be understood. But it is a mistake to look on the job as impossible. Clarity and simplicity are not incompatible, even if detail must suffer. For an author accustomed to writing for a scientific journal such as *Nature*, it is often difficult to see clearly what caveats and controls are required only by a professional audience and which are necessary for a balanced public presentation of a topic.

Confusion on this issue can easily result in obscurity or overinflated claims. Even if a particular topic may remain obscure, it is possible and useful to communicate whenever possible the enthusiasm and excitement that science engenders. What the research profession has forgotten is that enthusiasm can be infectious, and in such a way that everyone benefits. The alternative to exposing the scientific process is to make it appear mystical and beyond reproach. But people are inclined to expect miracles from those beyond reproach, and delivering on that expectation will often be a hopeless endeavour. □



Nobel prize for Japanese immunologist

London

A MIXTURE of surprise and delight greeted this week's award of the 1987 Nobel Prize in Physiology or Medicine to Susumu Tonegawa. The delight will be particularly marked in Japan, as Tonegawa, a Japanese national even though he has not worked there since graduating in 1963, is Japan's first Physiology-or-Medicine laureate. The surprise is that the Nobel assembly chose Tonegawa alone, whereas three weeks ago, for example, a Lasker prize for the same topic was shared by Tonegawa, Philip Leder and Lee Hood.

Tonegawa is cited for his discovery of "the genetic principle for generation of antibody diversity" — the problem of how the system can produce millions of different antibodies from, at the most, thousands of genes. In 1976, Tonegawa, with N. Hozumi (*Proc. natn. Acad. Sci. U.S.A.* 73, 3628; 1976) provided the first molecular genetic evidence in support of the favoured hypothesis of the time.

First suggested in 1965 by W.J. Dreyer and J.C. Bennett, the idea was that separate genes encode that part of an immunoglobulin (antibody) chain which is relatively constant in structure between antibodies and that part which is very variable, and that the genes are joined during lymphocyte maturation. The combinatorial possibilities of joining the variable and constant region genes would enable much diversity to be generated from relatively few genes.

"During the following two years", says the Nobel citation, "Tonegawa completely dominated this area of research." Others were quick to contribute, although sometimes not as quick as they might have been had not they been more constrained than Tonegawa by the then current problems of carrying out recombinant DNA work. Certainly by 1978 Philip Leder, Lee Hood and several others were contributing substantially to the understanding of the several mechanisms by which a diversity of antibodies is produced but few opportunities to be among those answering the most obvious questions escaped Tonegawa.

In 1981, Tonegawa moved to Massachusetts Institute of Technology, where he has concentrated on the molecular genetics of the T-cell receptor. After speaking on this subject at a *Nature* conference in 1986, he was pestered by the Japanese press as to why Japanese biologists do not win Nobel prizes, for which he blamed the Japanese system, and whether he would break the mould. They have their answer.

Peter Newmark

President's new AIDS commission in turmoil

- Three principal members resign
- Split over 'ideological differences'

Washington

THE credibility of the US presidential commission on AIDS (acquired immune deficiency syndrome) is in serious doubt in the wake of the sudden resignation last week of its chairman, vice-chairman and sole medical staff officer.

The resignations follow weeks of rumours that ideological differences between committee members were preventing progress towards the formulation of a national policy on AIDS. A first report is due in November and final recommendations next June.

The commission was in trouble from the moment it was set up just three months ago (see *Nature* 328, 373; 1987). Although it was given a comprehensive brief to make recommendations to President Ronald Reagan on medical, legal and social aspects of AIDS, critics claimed that its 13 members had been appointed more for their conservative views than for any expertise on medical issues. The commission proved unable to fill more than three of its 15 staff positions and its first executive director was asked to leave soon after appointment.

The chairman of the commission, W. Eugene Mayberry, chief executive officer of the Mayo Clinic, is unwilling to give a public explanation for his resignation. But the vice-chairman, Woodrow A. Myers, health commissioner of Indiana, who resigned shortly afterwards, has been more forthcoming, describing strong differences in ideological perspective and personality that prevented the commission from working efficiently. The commission's senior adviser on medical and research affairs, Franklin Cockerill of the Mayo Clinic, also resigned.

Attention is now focused on Dr Frank Lilly of the Albert Einstein Medical Center, who has been in the public eye not so much because he is the commission's only scientist but because he is its only homosexual. His appointment has been seen as President Reagan's one concession towards recognition of the difficulties of the group that has suffered most in the AIDS epidemic. After the resignations were announced, Lilly said that he was also considering resigning but would wait to see who the White House would appoint to replace departing members. The new chairman, already announced as retired Admiral James D. Watkins, a member of the commission, has attempted to persuade Lilly to stay on.

No early end is expected to the dissensions within the commission, chiefly because they reflect deep divisions within the administration and among the general public over what should be done to prevent the spread of AIDS. This week it was the turn of the Education Secretary, William J. Bennett, to join the fray with a new handbook, *AIDS and the Education of our Children — a Guide for Parents and Teachers*, which represents the Education



William J. Bennett: stresses moral education. Department's first official recommendations to US schools.

The handbook, which comes with the approval of the White House, toes the now familiar conservative line that, as Bennett puts it, "when it comes to AIDS, science and morality walk the same path, they teach the same thing".

While the values espoused by Bennett have not been criticized in themselves, doubts have been raised about the effectiveness of moral education as the first line of defence against AIDS. Bennett's handbook itself repeats statistics showing that, by the age of 19, three-quarters of all boys and almost two-thirds of all girls in the United States have been sexually active. More than half of all final-year high-school students have had experience with illegal drugs.

Given these figures, it is not surprising that Surgeon-General C. Everett Koop has taken a different approach. While he also accepts that monogamy will prevent the spread of AIDS, he has strongly recommended the use of condoms to the many young people who do engage in premarital sex. Bennett's handbook, in contrast, points out there is a danger that "promoting the use of condoms can suggest to teenagers that adults expect them to engage in sexual intercourse".

Alun Anderson

Critics denounce first genome map as premature

San Francisco

COLLABORATIVE Research, Inc. unveiled what it claims to be "the world's first genetic linkage map of the entire human genome" at the 7 October meeting of the American Society of Human Genetics, in San Diego. The Massachusetts company described its achievement as a scientific milestone, a valuable research tool and an avenue into the potentially lucrative production of diagnostic tests for common diseases. The map will appear in the 23 October issue of *Cell*.

Some of the company's critics, who characterize the map as simply one step in a continuing process, protest that the announcement is premature, amounting only to an act of "scientific merchandising". Orrie M. Friedman, company chairman, defended the move, saying that, as the map is ready for use, Collaborative was obliged to make it public to provide reassurance for its stockholders, credit for its scientists and an important resource for the scientific community.

The linkage map, with an average 10 centimorgan resolution, has emerged

from "islands of linkage" within the past several months, said Helen Donis-Keller, director of the project. The group has put 404 polymorphic sites in order on the genome, using DNA from 21 different families supplied by the Centre d'Etude du Polymorphisme Humain (CEPH) in France.

Collaborative's researchers sorted through more than 1,500 single-copy probes from a human DNA library. They found 250 revealing restriction fragment length polymorphisms (RFLPs) with heterozygosity of 40 per cent or more, making them highly informative markers likely to differ between individuals. The rest of the 404 markers came from individual chromosome libraries or were publicly available probes identified by others.

The polymorphic sites were ordered on the genome with the help of two computer program packages independently developed for the project, one by Collaborative scientist Philip Green and the other by MIT mathematician Eric Lander. The programs first compared the markers two by two, sorting them into linkage groups, then ordered the markers within those groups. In many cases the analysis was done with both programs to confirm the final map.

Linkage-groups were localized to specific chromosomes by hybrid panel analysis, in which a probe from a linkage group is hybridized against DNA from various hybrid cell lines, each of which contains an incomplete set of human chromosomes.

Donis-Keller said the map is by no means "complete" and needs more markers to improve resolution and to close gaps that remain on some chromosomes. Nevertheless, there is 95 per cent chance that any new probe will be linked to a known marker.

The map will increase the efficiency with which single-gene disorders can be localized, said Donis-Keller. With a handful of the most informative markers, a disease locus can be localized to a linkage group and then ordered among markers in the region.

In the interests of scientific openness, Collaborative will make its probes freely available for research purposes. It is, however, applying for patents on the probes, and stands to gain if an investigator uses a probe in a way that has commercial applications. In response to cries of commercialism, Friedman calls the plan the best compromise both to benefit the scientific community and to protect the company's shareholders.

The company's main goal in its five-year effort costing nearly \$12 million has been to produce a tool for the development of

diagnostic tests, but the market for tests for the relatively rare single-gene disorders is too small to provide the required return on the investment.

Stock analysts have thus criticized Collaborative on the grounds that its approach to genome mapping has been too grand and expensive, and have in particular criticized the company's failure to capture the market for various diagnostic tests, using publicly available probes whenever possible and concentrating on a few specific diseases.

Friedman defended the company's approach as the "cornerstone of a long-term strategy" to develop diagnostic tests for complex inherited diseases, such as heart disease, cancer, diabetes and psychiatric illness, that may be caused by two or more genes acting independently. He estimated that the company could increase its present \$13 million annual turnover (from the sale of various biomedical and pharmaceutical products) to several hundred million dollars with the introduction of tests for these diseases.

The genome map opens the way for the development of such tests, according to Green. With markers spaced across the whole genome, Green believes simultaneous computer-assisted analysis can link a disease trait to more than one locus. Without this approach, data from different families in which a disease is caused by different genes could cancel each other out. The present map, even with its gaps, should be sufficient to begin on such analysis, say the authors of the study.

Nevertheless, the gaps have led some critics to protest at Collaborative's announcement. Ray White, who is compiling a similar map at the Howard Hughes Medical Institute in Salt Lake City, said he considered publication of any map containing gaps to be premature, and added that he is publishing his map one chromosome at a time, as he develops complete linkage groups.

Collaborative officials point out the differences between the goals of the two groups. White is seeking a map with 1–2-centimorgan resolution, for the eventual cloning of disease genes, but Collaborative researchers have sought a map to help them begin on the genetic analysis of complex inherited diseases, on which they now feel ready to embark.

A map comparable to that of Collaborative Research is being developed by CEPH in Paris. That map will owe much to Ray White who has supplied more RFLPs than any other contributor to the collaborative project. Jean Dausset, who heads CEPH, says that its 10-centimorgan map will be finished in 1988. Two weeks ago, he announced plans to distribute the 150 or so probes that will define the map free of charge to any researcher on condition that they are not used commercially.

Marcia Barinaga

AIDS virus infects another lab worker

Washington

THE National Institutes of Health (NIH) last week reported the second case of a laboratory worker becoming infected with HIV, the virus causing AIDS (acquired immune deficiency syndrome), through occupational exposure (see *Nature* 329, 92; 1987). The infection evidently occurred when a worker at a facility under contract to NIH cut his hand while handling concentrated virus.

The worker's injury occurred toward the end of 1985, and the presence of the infection was first detected in May 1986 by an ELISA (enzyme-linked immunosorbent assay) test performed on a sample of the worker's blood. NIH officials and the worker in question were only recently informed of the test results.

Investigators will attempt to isolate and characterize the virus from the infected worker to see if it is the same as the one he or she was handling.

A team of safety investigators from NIH and the Centers for Disease Control in Atlanta is in the process of inspecting facilities licensed to handle concentrated quantities of HIV. It has so far concluded that current safety precautions are adequate to protect workers and the public so long as safety procedures are strictly observed.

Joseph Palca

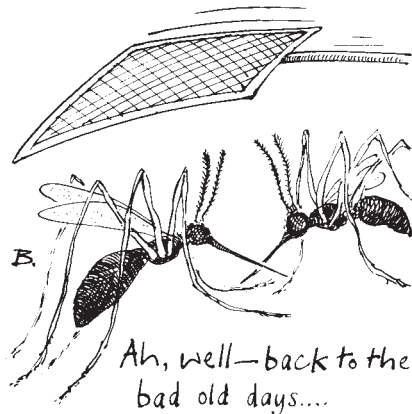
Indian health planners shun insecticides in disease control

New Delhi

AFTER having spent over 3,000 million dollars on insecticides in the past 25 years, India is convinced that it has lost its battle against mosquito-borne diseases. Now planners are testing an alternative control strategy based on environmental management with community participation — but no insecticides.

The Indian Council of Medical Research (ICMR) is spearheading the new integrated vector management (IVM) scheme, based on a combination of biological control and environmental measures such as cleaning clogged drains and filling stagnant pools. Results of the five demonstration projects conducted so far are encouraging and hold out hope that the approach may provide the final solution.

With half of its health budget spent on malaria prevention India is running the biggest anti-malaria programme in the



world. Insecticides and drugs lie at the heart of the programme but have brought with them a host of new problems. Bottled milk has been found containing 9.5 times the permissible limit of DDT, and samples of human milk collected in Punjab contained 21 times the acceptable limits of various pesticides. At the same time the development of insecticide resistance has made new and more expensive imports necessary.

The new strategy aims "to strike a balance between man, mosquitos and the environment in such a way that the vector population is stabilized at a very low level at which disease transmission ceases to occur", says Dr P.K. Rajagopalan, director of the Vector Control Research Centre (VCRC) in Pondicherry, one of the ICMR institutes. VCRC applied its new strategy in the filaria-endemic coastal township of Pondicherry, where mosquitos were breeding in thousands of cesspits, ponds and wells, a large swamp and 100 km of open drains. First, VCRC launched a massive health-education campaign aimed at enlisting public cooperation in

cleaning up the town. Within two years 15,000 m² of surface water were eliminated and wells stocked with larvae eating fish. The swamp was converted into a park. In 1986, when the project ended, there were only three new cases of filariasis in the whole of Pondicherry and the number of mosquito bites, according to a VCRC report, "dropped by 90 per cent".

Recently ICMR applied its new app-

roach to an industrial township (population 45,000) in Hardwar on the banks of the Ganges in Uttar Pradesh. In its interim report ICMR said that within 18 months of launching the project, mosquito densities fell by half while cases of malaria dropped from 250 to 32. That improvement was achieved simply by filling ditches with fly-ash (an industrial by-product), cleaning drains and covering overhead tanks. The 30,000 dollars spent on environmental control was just five per cent of what industry would otherwise have lost from paying for sick leave and hospitalization of workers with malaria. **K.S. Jayaraman**

Training of operators needed for nuclear power station safety

London

A SOVIET automation expert has described the training of nuclear power station personnel as "amateurish". Writing in the daily *Sotsialisticheskaya Industriya*, R. Tsipyura, described as a department head at the Promavtomatika scientific and production association, noted that although the responsibility of operating a power station control board is analogous to that of an airline pilot, neither the USSR Ministry of Power and Electrification nor the USSR Ministry of Nuclear Power has a training scheme similar to those used in aviation. For nuclear stations, he said, training is carried out by on-the-job and understudy methods.

Recently, he notes, there have been some changes: the Novovoronezh training and instruction centre now has simulators for the VVER-400 and VVER-1000 reactors, and this year a "technological process control station", providing training for RBMK (Chernobyl-type) reactors will come into operation at the Smolensk nuclear power station.

In spite of these apparent changes, however, he said, the two relevant ministries have no proper training system, and "incredible" things sometimes happen; such as people who have no real connection with the operator's job being sent for training, simply to "fulfil the training plan" in gross figures.

Simulator training was one of the measures promised by the Soviet delegation at the post-Chernobyl accident review conference in August 1986. Had the Chernobyl crew been trained in simulated accidents, the Soviet team admitted, they would not have overridden the strict operational rules so lightly.

But bad operator procedures are not the monopoly of the Soviet nuclear industry — Tsipyura noted that at the Ekibastuz Hydro-storage power station during 1981–1984 there were 366 accidents due exclusively to operator error. But when the power-production sector was split, by

hiving off nuclear power into a separate ministry, the responsibility for simulator training was likewise divided. The present simulator training units have been developed by individual power stations. Tsipyura suggested that efforts should rather be coordinated into a network of "inter-department science and production collectives" subordinate to some inter-departmental body such as the State Committee for Science and Technology, or the USSR Council of Ministers' Bureau for the Fuel and Power Complex.

"Interdepartmental collectives" is one of the catch-phrases of *perestroika* (restructuring). When first urged by Mr Gorbachev in June 1985, they were intended to overcome the bureaucratic and inter-sector delays bedevilling the Soviet economy. So far, however, anecdotal evidence suggests that they are encountering bureaucratic problems of their own. Tsipyura may therefore be over-optimistic in suggesting that such a network would necessarily resolve the situation. **Vera Rich**

Tropical fruits

London

Two funding organizations have joined forces to help finance the twinning of institutes in developed and developing countries that share a research interest in tropical diseases. Up to \$2 million a year will be provided to worthy partners under the new scheme, which is financed by the Rockefeller Foundation and the UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Disease, under its new director, Tore Godal. Value for money will be judged by the scientific productivity of the joint venture and the extent to which it promotes intellectual self-reliance of the weaker partner. But there will be maximum flexibility on what the money is spent on, says Godal. **P.N.**

Pasteur Institute celebrates centenary — and so do NIH

Paris

THE renowned Pasteur Institute in Paris opened its centenary celebrations last week with a five-day international symposium on molecular biology and infectious diseases. More than 300 researchers and physicians, including eight Nobel prizewinners, attended the conference, which was opened by the French President, François Mitterrand, and closed by the Prime Minister, Jacques Chirac.

Other events associated with the celebrations — estimated to have cost its sponsors around FF10 million (\$1.6 million) — include a retrospective exhibition at the institute, a 50-minute television documentary film, a photographic exhibition at the national science and industry museum and the creation of a permanent historical exhibition at the Pasteur station on the Paris metro.

The four themes of the symposium — viral diseases, bacterial diseases, parasitic diseases and host defence mechanisms — reflected developments over the past century in microbiology as a whole and in the work of the institute in particular.

Many 'Pastorians' found a symbol of the century's progress in microbiology in the presentation at the symposium, by its youngest speaker, Noël Tordo, of the complete genome of the rabies virus. The institute was founded in 1887 after a fund-raising drive provided Louis Pasteur with an institution in which to develop the rabies vaccine he had successfully demonstrated two years before.

Regulations covering the release of recombinant microorganisms could, however, preclude the use of new rabies vaccines being tested at the institute as part of this campaign.

The Pasteur Institute has remained an independent foundation, despite a difficult time in the 1970s when the government was called upon to provide extra funds. Since 1976, 47 per cent of the institute's FF500 million annual budget comes from the state. In addition, the state pays for salaries of 140 of the institute's 500 staff researchers attached to the Centre National de la Recherche Scientifique (CNRS). Thirteen of the institute's research units (include Luc Montagnier's viral oncology unit) are associated with the CNRS.

More than half the institute's budget is made up of a combination of fund-raising (14 per cent), royalties from its industrial subsidiaries, Pasteur Vaccins and Diagnostics Pasteur (10 per cent), and from its own resources (29 per cent). The creation of a separate commercial infrastructure to exploit the institute's discoveries began in the 1970s and replaced the previous 'cot-

tage industry' approach to vaccine preparation in institute laboratories.

In his closing address last Friday, Chirac tried hard to find a way to make it seem as if his government cares for fundamental research — most public sector budgets for 1988 will either be cut or, effectively, remain static relative to last year, although medical research is less affected (see *Nature* 329, 380; 1987). Chirac referred to the FF100 million additional grant recently released for AIDS (acquired immune deficiency syndrome) research — part of which will go to the institute — but had to look back to the era of President Charles de Gaulle for other evidence of Republican legacies to medical research.

Celebrations will continue into Novem-

ber and hopes are that additional funds will be attracted as a result of the high public profile of the institute. **Peter Coles**

● On the other side of the Atlantic, the National Institutes of Health (NIH) are in the midst of the concluding celebrations to their own centennial year. NIH began rather more modestly than the Pasteur Institute as a one-room Hygienic Laboratory on Staten Island, New York, but have grown into a massive complex of laboratories centred on the campus at Bethesda, Maryland, and employing a total of 3,500 scientists. Their \$6 million budget supports the research of another 20,000 scientists at 1,300 universities throughout the United States. Between them, NIH grant-holders have won close to 60 Nobel prizes.

The history and future of NIH will be the subject of two special articles in next week's *Nature* from former NIH directors Robert Q. Marston and Donald S. Fredrickson. □

British government crushes hopes for extra space funds

London

THE mood of cautious optimism within the British space community, which believed that there was a real chance of the government increasing its investment in space research before next month's ministerial meeting of the European Space Agency (ESA), has been destroyed by the announcement by Mr Kenneth Clarke, the Minister for Trade and Industry, that no more money will be forthcoming.

Speaking on television at the weekend, Clarke described ESA as a "hugely expensive club" with over-ambitious pro-



Kenneth Clarke — no more money.

grammes. The government would not be increasing its £112 million space budget, three-quarters of which is spent through ESA. When ESA ministers meet on 9 and 10 November in The Hague, they are expected to ratify a major infrastructure programme for the next decade, requiring at least a doubling of the present budget.

The timing of Clarke's announcement has angered and puzzled the space community. It had been widely understood

that the government's newly created Advisory Council on Science and Technology (ACOST) would review Britain's space programme before the ESA meeting and make recommendations to the government. After the first meeting of ACOST on 29 September, its chairman, Sir Francis Tombs, said that "preliminary consideration" had been given to national priorities for space research. Tombs said that a "round-table discussion" would be held between ACOST and representatives of the space community. It was widely assumed that this meeting, scheduled for later this month, would enable ACOST to decide if Britain should increase its spending on space, and advise the government accordingly.

Since the prime minister, Mrs Margaret Thatcher, announced in the House of Commons on 23 July that the government would not increase its space budget "for the time being" (see *Nature* 328, 467; 1987), Britain's space community has been in turmoil.

Roy Gibson, BNSC's director-general, resigned on 4 August. Later that month the government bowed to pressure from ESA and industry by producing an extra £4 million to enable British companies to continue pre-development work on ESA projects until the November meeting. Given this encouraging sign, the British space community refused to accept that all was lost, pinning its hopes on ACOST.

The space community seems to have been prompted by its own enthusiasm into believing that the government was on the verge of reversing its earlier decision. Mr Clarke has put paid to that.

Simon Hadlington

Hughes Institute gives US science education a boost

Washington

WITH \$500 million to spend on grants as a result of its settlement with the Internal Revenue Service, the Howard Hughes Medical Institute is moving into new areas of support for science education. Awards were announced last week totalling \$45 million, intended largely to complement the institute's traditional support for basic research by helping to ensure the supply of talent for biomedical research.

For non-US citizens, the new programme that will attract most attention is the provision of doctoral fellowships — 60 in 1988 and more in succeeding years until 300 are available — which are open to all graduate students, regardless of nationality. The 3–5-year fellowships are generous, with \$23,000 a year provided for living expenses and fees. Those interested will have to hurry: applications for 1988 must be in by 13 November. The job of selecting from among the applicants will be contracted to the National Academy of Sciences.

By far the largest share of the institute's new spending will go to an undergraduate biological sciences "education initiative". With \$30 million on offer, the institute will at one stroke become the biggest player in

the field, surpassing spending by the National Science Foundation on improving undergraduate science education.

Ninety-four undergraduate institutions, almost 20 per cent of them with a history of high enrolments from racial minorities, have been invited to apply for individual awards in the range from \$500,000 to \$2 million. Curiously, it is the colleges that are already doing well at sending on students to biomedical research careers that have been invited to apply, rather than those that are doing badly — presumably on the grounds that the best are most able to do better.

Increasing the very low number of minority students who pursue a career in science is one explicit goal of the programme, but ideas on how biomedical education might be improved will be sought in imaginative grant proposals

rather than dictated by the institute.

Two laboratories will also receive large awards to help educational programmes. Seven million dollars goes to the Cold Spring Harbor Laboratory for expansion of education in the neurosciences. The three-year grant will help to establish an \$11-million facility providing year-round training in advanced areas of neurobiology and provide for a permanent research programme with the emphasis on molecular approaches to neural development. Two million dollars goes to the Jackson Laboratory to improve training facilities and to increase the capacity of facilities for storing transgenic mouse embryos.

A further \$5 million will be provided at the Institute of Medicine for the development of programmes that promote public understanding of science and studies in health sciences policy. The latter are expected to help the institute to identify areas of research where development could be accelerated and where basic research discoveries could be put to use in medicine.

Alun Anderson

Learned societies hit back at UK government education plans

London

PROPOSALS to create a three-tier system of scientific research and teaching in Britain's higher-education institutions have been condemned by representatives of 16 scientific societies and institutions who met in London earlier this month to discuss the document *A Strategy for the Science Base*, produced in July by the Advisory Board for the Research Councils (ABRC) at the request of the government (see *Nature* 328, 280; 1987).

The document's most contentious proposal is to categorize institutions into three types, with most research concentrated in about 15 'type R' institutions; a limited amount of specialized research and teaching in type X institutions; and teaching-only ('type T') institutions. At last week's meeting, arranged by the Biochemical Society, the proposal was rejected outright. It was concluded that "the T concept is unrealistic and should be eliminated forthwith", that a combination of R- and X-type institutions would be "difficult to implement although perhaps could be acceptable". The preferred option would be for a system of type X institutions only, which should be positively developed by the ABRC and the University Grants Committee "in respect of selectivity, collaboration, concentration of scientific effort and funding, and interaction with industry".

The concept of interdisciplinary university research centres (URCs) was also decried. The Science and Engineering Research Council has already invited bids

from universities to host such centres in various subjects. It was concluded at the meeting that there is no evidence that the concept is viable and that in effect the URCs would be heavily funded establishments that would be difficult to dismantle were they to be ineffective or unproductive, while at the same time "starving individuals in other institutions of support for their research". The ABRC document makes "inadequate allowance for the fostering of creativity and innovation."

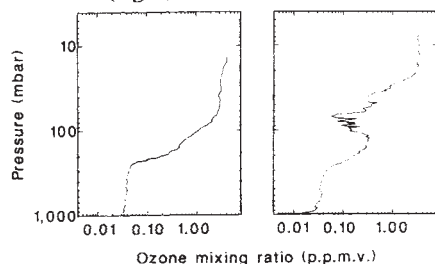
A further criticism of ABRC is that it has neglected to consider research teaching hospitals and medical schools, particularly the question of the type of institution to which the independent medical schools in London would be attached. There is also concern about the lack of thought given to distribution of charitable funding, should the three-tier system be adopted. "Positive measures" are needed from the government to promote constructive interaction with industry, including the introduction of tax-incentive schemes.

The government says it will not make a decision on the ABRC proposals until it has considered the responses to the document, for which the deadline is 31 October. Nearly all of Britain's scientific societies and institutions were represented at the meeting. They will submit individual responses, and next month ABRC will meet with representatives of the societies and the Committee of Vice Chancellors and Principals for more discussion.

Simon Hadlington

Ozone depletion over Antarctica

THE depletion of ozone over Antarctica this year is the most severe yet (see *Nature* 329, 473; 1987). The stark contrast between profiles of the ozone layer at Halley Bay (76°S, 27°W) in mid-August (left) and early October (right) is shown. The column



amounts on these days were 296 matm-cm and 128 matm-cm respectively (provisional values). The heart of the layer has been removed. At 70 mbar (about 16.5 km), 97.5 per cent of the ozone has been destroyed. There are now two weak layers, with peak amounts at about 160 mbar (12 km) and 22 mbar (23 km). In the core of the polar stratospheric vortex (100 to 30 mbar, 15 to 22 km), ozone amount has fallen by practically an order of magnitude, from 130 to 15 matm-cm.

J.C. Farman & B.G. Gardiner

Reorganization of biology departments on Berkeley campus

Berkeley, California

FACULTY and graduate student recruitment is up as the University of California at Berkeley begins an unprecedented reorganization of its biology departments. Ten departments will be dismantled and two created, while two new biology buildings will be put up and a third completely renovated.

Fund-raising for the \$150-million building project began last year (see *Nature* 321, 7: 1986) and has raised \$19 million so far, 122 per cent of the level expected for this point in the drive. In a two-to-one matching agreement, the state has pledged two-thirds of the money needed for the project. The university will raise the other \$47 million as part of a two-year campus drive seeking \$320 million for six major projects.

The reorganization, which will redistribute but not enlarge the biology faculty, was conceived in the early 1980s, after both internal and external review committees concluded that biology at Berkeley was slipping in the national rankings because of outdated laboratories and archaic department divisions.

"Our departments are not congruent with modern biological sciences", said developmental biologist Fred Wilt. For example, developmental biologists are

scattered across the campus in five or six different departments, making it difficult for them to act as a group to recruit graduate students.

Such problems are likely to disappear when the new scheme sorts the faculty into departments that accurately reflect their research. Laboratories will be spatially organized into "affinity groups" that share common interests and techniques, to optimize intellectual exchange as well as the use of equipment and facilities.

The biggest reshuffling will be the creation of one large cell and molecular biology department, with 91 members in six divisions: genetics, neurobiology, biochemistry and molecular biology, biophysics, cell and developmental biology and immunology. Faculty may be affiliated with more than one division, and can change divisions as their research interests shift. Graduate students in the department may move among divisions.

The new biology department will have 36 faculty members in the areas of evolutionary, ecological and organismal biology. The 20-member plant biology department will be an expansion of the present molecular plant biology department, with members coming from the dissolved botany and genetics departments as well.

Many of the more applied departments,

such as plant pathology and entomology, will remain intact, although some members will join affinity groups containing members of new departments.

Such a major metamorphosis must take place in stages, said Daniel E. Koshland Jr, chairman of the chancellor's committee overseeing the project. The old departments are still operating, but a "shadow government" of new department and division chairmen is already in place, formulating curricula and preparing to take over within a year.

By that time, all the laboratories will be occupied in the first of the two new buildings, the Life Sciences Annex, and the second building will be under construction, with completion planned for 1989.

Renovation of the overcrowded and unsound 55-year-old Life Sciences Building is the final and most ambitious leg of the project, scheduled to take three years, from 1989 to 1992. Vacant faculty positions may not be filled until the renovation is complete, while laboratory relocation will be a nuisance, but several faculty members say that the intended temporary space will be far superior to what they have now.

Marcia Barinaga

India aims at own supercomputer

New Delhi & Bangalore

INDIA will attempt to design and build a parallel processing computing system in three years in a \$50 million project approved by Prime Minister Rajiv Gandhi.

A new organization, the Centre for the Development of Advanced Computer Technology (CDACT) has been created in Pune, near Bombay. The goal is to develop a parallel computing system with a minimum rating of 50 megaflops and capabilities equivalent to a bottom-level commercial supercomputer.

The new mission implements a recommendation of the Prime Minister's scientific advisory committee headed by Professor C. N. R. Rao. Increasing demands for high-speed computing by Indian scientists, coupled with their failure to obtain a supercomputer of their choice from abroad, have prompted the government to develop its own technology.

CDACT will be answerable to the Department of Electronics but will have considerable autonomy. A spokesman for the department says that the time is ripe for India to enter the field of parallel processing technology because the methods and architecture being developed elsewhere are in their early stages. Assistance will come from many other agencies, including the Indian Institute of Science, the National Aeronautic Laboratory and the Tata Institute of Fundamental Research.

K.S. Jayaraman & Radhakrishna Rao

Confronting natural disasters

Washington

WHEN the largest earthquake in recorded US history hit Alaska in 1964, at magnitude 8.4–8.6 on the Richter scale, this was the result on Anchorage's Fourth Avenue as streets buckled and buildings sank by as much as 4 m. Since then, enormous advances have been made in engineering structures that will reduce casualties and damage from natural disasters, including not just earthquakes but also landslides, tidal waves, windstorms, floods, volcanic eruptions and wildfires.

Prediction of hazards has also come into its own with radar making it possible to track storms more accurately. Some large-scale international projects have also proved successful, such as the Pacific Tsunami Warning Center which provides advance warning of tidal waves. But much of what is known about hazard reduction has as yet been applied only in a few parts of the world.

That is the major conclusion of a new report* from the National Academy of Sciences which stresses that "there is a vast opportunity to advance our know-

ledge of hazard reduction if we pool resources". If already available information can reach local officials there are good prospects for improving on a record which has left 2.8 million people dead



Buckled streets

from natural disasters in the past 20 years.

Towards that goal, the report recommends that an International Decade for Natural Hazard Reduction (1990–2000) be established with the full backing of the United Nations to facilitate cooperative research, technology transfer and training. Leadership could be provided by a strong US programme that, the report says, could "cut impacts of natural hazards at least 50 per cent by the year 2000".

Alun Anderson

*Confronting Natural Disasters (National Academy Press, 1987).

European research centres may have to pay their way

London

PROPOSALS to introduce major changes to the management and work of the European Community's Joint Research Centre (JRC) — which runs laboratories at four separate sites — are due to be announced later this month, officials in Brussels have confirmed. The JRC's role in the Community's research programme has come under close scrutiny during the past year, particularly throughout the prolonged deliberations over the size of the budget for the five-year research and development 'Framework' programme. Shortly after the programme was officially ratified at the end of last month (see *Nature* 329, 381; 1987), Karl-Heinz Narjes, the commissioner with responsibility for technology, told a private meeting of research ministers that reform of the JRC would be announced imminently, with particular emphasis on a less rigid management structure, a review of the inflexible 'multi-annual' system of research planning and greater collaboration with industry, both to earn money by contract work and through 'industrial clubs' to ensure a greater degree of 'relevance' in the research. It is understood that the Commission wants to see the JRC earning around 40 per cent of its own income by 1990.

Reform of the JRC is seen as a way of appeasing critics of the Community's research programme, particularly Britain and West Germany, the two countries largely responsible for the delay in approving the Framework budget. But even the more sympathetic member states agree that a shake-up is due. The JRC was set up under the Euratom treaty as a nuclear research centre, becoming operational in 1960–61. Most of the expertise is in nuclear matters, although the scope of its work has expanded more recently to include such fields as remote sensing and the study and protection of the environment. The JRC comprises four separate establishments, at Ispra in Italy, Petten in The Netherlands, Geel in Belgium and Karlsruhe in West Germany.

The grounds for a reorganization were laid last year with the publication of the JRC's mid-term evaluation report by the Commission's scientific council. That report condemned the multiannual programme system as imposing "stifling rigidity". Decision-making procedures were "complex and ponderous" and although most of the research activities were of "generally good quality", the JRC needed to increase its links with outside institutions, including industry.

A report from an independent panel of prominent European industrialists was

also commissioned. The report, presented to the European Commission earlier this year, remains unpublished but is known to have been critical of the management and research priorities of the JRC, particularly at Ispra which accounts for three-quarters of the total JRC budget and staff.

This year the JRC will receive approximately 200 million European Currency Units (1 ECU = £0.69), of which 21

million ECU will come from sources other than the JRC programme budget (most of it is accounted for by 14.5 million ECU from the West German and Dutch governments' contribution to the high-flux reactor project at Petten). While officials maintain that much of the criticism of the JRC is unfair, they concede that the JRC's outside earnings are 'marginal' and that there is considerable scope for improvement. The total staffing level of 2,260 is likely to remain the same, although the need for new blood has been stressed and is likely to be reflected in any new recruitment policy. **Simon Hadlington**

Permanent assistant's frustration leads to professor's murder

Tokyo

SUPERCONDUCTORS and murder? When the body of a Hiroshima University professor who specialized in superconductor research was found on the floor of his office in late July with knife wounds in the back, it was speculated that intense rivalry in the race to develop high-temperature superconductors lay behind the incident. But although the new superconductors played an indirect role, it now appears that the murder had more to do with appointment practices and the problem of 'permanent assistants' in Japan's national universities.

Dr Hiroshi Suemitsu, an assistant in elementary particle physics in Hiroshima University's Faculty of Integrated Arts and Sciences, has confessed to killing his professor, Tetsuhiko Okamoto, dean of the faculty, for passing him over for promotion. At a faculty meeting in July, it was decided that two posts to be vacated by associate professors of elementary particle physics in February next year will be filled by material scientists, including a specialist in the new superconducting ceramics. For Suemitsu this was the final straw in his 17-year struggle to climb above the first rung of the academic ladder.

Joshu, or assistant in Japanese, literally means 'helping hand', and some professors interpret the term that way, assigning all sorts of menial tasks to their assistants. But the post carries no teaching responsibilities, and, apart from professional chores, assistants are free to carry on research, albeit with very limited funds.

As civil servants, *joshu* are entitled to lifelong employment, and, although the salary is below the national average, some of Japan's assistants seem content with their lot. As one Tokyo University assistant put it, "there are essentially two species of *joshu*: those who see the post as a stepping stone to an associate professorship and beyond, and those who seem to have settled for life". But even the former

species often get stuck on the first stone — many of the approximately 17,500 assistants in Japan's national universities are in their 40s. And the problem of 'permanent assistants' is a major headache for the Ministry of Education, Science and Culture, their employer.

From his confession, it is clear that Suemitsu was seeking promotion. He graduated from Hiroshima University in March 1966 and was appointed as an assistant in 1970 before completing his doctorate in 1972. At that time, Hiroshima University faculty were drawn almost exclusively from the ranks of the university's own graduates, a situation that still prevails in many of Japan's national universities. But following the student riots of the early 1970s, Hiroshima University underwent a major reorganization and in 1974 Suemitsu was placed in the new Faculty of Integrated Arts and Sciences, which has a more open and flexible appointment system. After the reorganization, the university received a flood of appointees from other universities such as Tokyo and Nagoya. So for Hiroshima graduates such as Suemitsu, competition for faculty places became even fiercer.

Moves are being made to try and cope with the problem of permanent assistants. For example, at Tokyo University's Institute of Solid State Physics, there is an understanding that assistants should move on, either up the academic ladder or out of the institute, after about five years, according to Shuntaro Watanabe, an associate professor at the institute. But the problem is deep-rooted and can be traced to the expansion of the universities in the 1960s which created a huge population of PhD holders with no posts for them to fill (see *Nature* 309, 659; 1984). A decision to expand the universities seems most unlikely in the current climate of government financial restraint and as long as faculty enjoy lifelong employment, the assistants can only sit and wait for their superiors to retire. **David Swinbanks**

Comecon joint research under strain because of complexity

London

A remarkably frank account of impediments encountered by Comecon countries in their plan for joint research and development is emerging in the wake of last month's meeting of the party secretaries in Sofia. According to *Pravda*, one of the most serious difficulties is that European Comecon countries have set up separate research and development facilities, especially in microelectronics.

This admission is significant because the objective of the "comprehensive programme for scientific and technical progress", launched in December 1985, was to pool technical resources in five key areas (microelectronics, automation, nuclear power, new structural materials and biotechnology).

Earlier in the year, Comecon secretary Vyacheslav Sychev was saying that the programme had started slowly because of its complexity (more than 1,500 research institutes are involved) and the legal and cultural differences between Comecon countries. But even before the launching of the programme, Comecon has been stressing the benefits of planned diversification of research and the avoidance of duplication.

It has been known for some time that particular countries have been unhappy with some conclusions of the "fraternal discussions" by which decisions on the sharing of the burden of research and development are supposed to be made.

Bulgaria was thus given special help to developing its computer industry, while East Germany was dismayed at having to switch to new Comecon standards, threatening to make it diverge from West German standards (and was consoled by being made the host for the Comecon Standards Commission, allowing it to limit the divergence).

Cooperation with Comecon is most familiarly represented by joint projects in space and nuclear research, although the high magnetic field/low temperature institute at Wroclaw may be a more interesting example of joint research. But specific and often bilateral collaboration in industry and agriculture have not always achieved their goals.

Thus Dr Stoyan Markov, president of the Bulgarian State Committee for Science and Technology, has said in an interview in the Slovak *Pravda* that many cooperation projects have been ill-chosen, have involved obsolete technology and have brought no real benefit to those concerned.

Comecon cooperation with its three less developed members (Cuba, Mongolia and Vietnam) raises special problems taken up

at Sofia. Mongolia has been helped in the past by geological surveys, followed by strip-mining of the ore located, but is now asking for help with biotechnology.

Cuba has been helped chiefly with

resource exploitation and electric power generation, and while the industrial members of Comecon have helped with the construction of more than 300 manufacturing plants in Vietnam (some of them enterprises destroyed in South Vietnam during the civil war), there are now complaints that some of the machinery supplied is unsuitable for the tropical conditions.

Vera Rich

West German ructions over US surrogacy company

Munich

A US commercial surrogate motherhood company is on a collision course with the West German authorities after opening an office in Frankfurt on 1 October.

The company, United Families International (UFI), was given notice on 9 October by Frankfurt city officials that it is in violation of West Germany's adoption law and is subject to fines of up to DM 50,000.

But UFI representative Franklin Torch disagrees, saying that "West Germany is powerless to challenge" his company, which signs no contracts with the would-be parents but rather refers them to its parent company in the United States, the Infertility Center of Michigan (ICM).

ICM is the company whose New York office organized the adoption of 'Baby M' in New Jersey in 1986. The New Jersey courts set a US precedent when they awarded custody of Baby M to the biological father and his wife despite the protests of the surrogate mother who bore the child. (That ruling has been appealed against.) UFI is believed to be the first surrogate motherhood company to open an office in Western Europe.

The West German Ministry of Justice spoke out strongly against the offer of "mothers for hire" to childless West German couples. Spokesman Jürgen Schmid called it "the purest form of trafficking in human beings" and predicted that UFI would never win in a West German court. He said that only state-run or state-approved adoption agencies are allowed to operate in West Germany. Furthermore, West German adoption law would award a baby borne to a surrogate mother to the surrogate, even if she had signed a contract to turn the baby over to the biological father and his wife. Such contracts are considered null and void in West Germany.

Changes in the West German adoption laws are planned to protect the rights of the surrogate mother even more thoroughly. If approved by the Bundestag, the changes would give the baby to the surrogate even when both sperm and egg, fertilized *in vitro*, come from the would-be adoptive parents.

UFI reports that it has received several requests for information. The whole process, including medical and legal fees, normally costs \$30,000. ICM has served as the go-between in the adoption of 184 surrogate babies so far and did \$700,000-worth of business in 1986.

Although ICM customers have come from as far away as France, Italy, Israel, Greece and Australia, all of the surrogate mothers so far have been from the United States or Canada. Each surrogate mother receives payment of \$10,000 for bearing a child.

Torch reported that "three or four" West German couples have already adopted babies through ICM in the United States. Apparently these cases have not come to the attention of the Justice Ministry, whose spokesman said West Germany would categorically deny adoption papers to couples so "pathological" as to want to adopt a baby in this manner.

ICM's international business may expand even if West German authorities succeed in closing UFI, which will surely not go down without a fight. The publicity surrounding UFI's opening may have been enough free advertising to last for years. And it is questionable whether West German authorities could determine that newly adopted babies had been born by this procedure, as it is possible for the adoptive parents to obtain a second birth certificate for the baby citing the wife as the baby's mother.

Steven Dickman

Mars trip far off

London

No manned trip to Mars can be expected before the turn of the century, according to Cosmonaut Aleksandr Serebrov.

Addressing a press conference in Budapest to mark the 30th anniversary of the Soviet satellite Sputnik-1, he said that no "official decision" about a future Mars mission had yet been made, but that tentative longer-range plans were being drawn up for an international mission. Short-range plans, to the end of the century, would "create the conditions" for a manned flight to Mars.

V.R.

National curriculum

SIR—Your leading article (*Nature* 328, 356; 1987) on the British government's proposal for a national curriculum, in primary schools as well as secondary schools, failed to address all the many difficulties in the proposal.

While a set curriculum in all schools might be advantageous to the small mobile fraction of the school population (usually the children of the upper echelons of management) most children within an educational authority are static.

The attainment tests at 7, 11, 14 and 16 years, proposed but not yet defined, are fraught with difficulties. The young are not uniform sponges that can imbibe a given amount of water and release it on being squeezed, but form a mixed population with a wide distribution of ability and rate of educational development. To set a national 'norm' for attainment will divide the school population into two classes — those above the 'norm' (pass) and those below (fail). And would the evaluation of the 'norm' include the fraction of the school population at present considered 'educationally subnormal'?

Much has been written about the beneficial effects of parental involvement and encouragement and the home environment (availability of books) on a child's academic attainment. This government's contention that teachers are responsible for "diminished expectations in young people" is not necessarily correct because the home environment does determine, to a large extent, a young person's expectations. Poverty, because of low (or no) income and poor job prospects, despite educational success, is bound to influence adversely a young person's expectations and induce a feeling of hopelessness which no amount of 'teaching' could overcome.

So is it valid to compare teaching in schools that draw from different catchment areas on the basis of their attainment rates? In multistream schools, will teachers of the 'high flyers' be rewarded while the teacher of the 'backward class' stays bottom of the salary scale?

Teaching towards so many attainment tests, if success/failure is to be taken as a measure of teaching ability, is guaranteed to stultify education and reduce it simply to 'learning by rote'.

Finally, your comments about Welsh suggest that you have little appreciation or knowledge of the language or the people. The Welsh have a cultural background at least equal to and which predates 'English'. If the British government were to observe EEC directives, it should actively encourage the preservation of the Welsh language and culture in the principality.

IOLO AB ITHEL DAVIES

Department of Pharmacy,
Queen's University,
97 Lisburn Road, Belfast BT9 7BL, UK

Plymouth brethren

SIR—Contrary to your report (*Nature* 329, 96; 1987) on the proposed merger between the two marine laboratories in Plymouth, the Laboratory of the Marine Biological Association and the NERC Institute for Marine Environmental Research, it is not a "stated goal" of this proposal that the merger should effect "cost savings". The stated goals, to quote from the draft Agreement, are "to formalize the collaboration between the two laboratories", "to harness their scientific strengths and reputations into one centre of excellence" and "to safeguard the future of research and related activities in Plymouth".

These objectives are primarily scientific and designed to enhance the research in Plymouth at a time of considerable scientific challenge in the marine sciences which demand closer collaboration and interdependence between the traditional disciplines of physical, chemical and biological oceanography.

Neither partner in this merger is unaware of the financial difficulties facing the marine sciences within NERC and elsewhere; we face these difficulties whether as one laboratory or two. But we also wish to respond to the longer-term scientific challenge, by merging our experience, our expertise and our enthusiasm. The true motives for the merger are to be found here, not in response to the immediate and, we hope, transient financial climate.

BRIAN L. BAYNE
(Director)

Institute for Marine Environmental
Research,
Prospect Place, The Hoe,
Plymouth PL1 3DH, UK

Yeast the model

SIR—Anne McLaren (*Nature* 328, 10; 1987) appears to have a very personal bias against the word 'embryo', despite its ancient origins and readily translatable meaning (the growing human, in this case), to describe the product of the union of male and female gametes. Her failure to see any significance in the unicellular human embryo may be blinding her to the reality present and to the significance of the 'kingdom' of unicellular organisms. As E.B. Wilson has indicated: "Long ago it became evident that the key to every biological problem must finally be sought in the cell, for every living organism is, or at some time has been, a cell".

It is now almost 30 years since the basic animal research work of 'in-vitro fertilization' (IVF) was completed and 25 years since the first attempts at human IVF². Despite the great promises of spin-off from IVF research into the field of genetic diseases, no therapies have become available.

But Ira Herskowitz, geneticist and a

recent recipient of a MacArthur Fellowship (see *Nature* 327, 680; 1987), has indicated that a unicellular organism, *Saccharomyces cerevisiae*, could be the most important model for answering the big questions of cellular biology: "If you want to understand a process at the molecular level, the best — and in some cases the only — place to start with is yeast"³.

The sooner Anne McLaren and others get off the unethical 'bandwagon' of human embryo research and onto the ethically suitable model of yeast research, the sooner will the biological problems of infertility and genetic diseases be solved.

PATRICK W. GILL

Neath General Hospital,
Neath, Glamorgan SA11 2LQ, UK

1. Wilson E.B. *The Cell in Development and Heredity* 3rd edn (Macmillan, New York, 1925).
2. Biggers J.D. *New Engl. J. Med.* 304, 336–342 (1981).
3. *Discover* 71–83 (November, 1986).

Schrödinger v. Marx

SIR—M.F. Perutz, in his evaluation¹ (or devaluation, to be precise) of Schrödinger's *What is Life?*² has committed the classical error of the description of an elephant by three blind men. Each of his statements is scientifically correct but, alas, the summation is wildly off-target. The analysis of what is wrong in *What is Life?* is surely of historico-scientific value, but then to denigrate the great book as a scientific advertisement of Timoféeff, Zimmer and Delbrück's paper³ is a bit too preposterous.

So far as science is concerned, the concepts of open systems and irreversible thermodynamics, to name but two, have derived from this little book. But surely, when one judges the merit of a book decades after its publication, what matters should be the role played by the book in the growth of subject that the book originally dealt with. In that respect, *What is Life?* has stood the test of time by being the grand initiator of the present integrated view of physics and biology.

It is interesting to compare *What is Life?* with Darwin's *Origin of Species* and Marx's *Das Kapital*. To judge from the present knowledge of molecular biology (as espoused by Kimura in his *Neutral Theory of Evolution*⁴ and the economic and political conduct of socialist countries (*à la glasnost* and *perestroika*), both books would no doubt be found "highly erroneous", but this judgement would not prevent them from being among the greatest books ever written.

SANTANU DATTA

Department of Clinical Genetics,
Karolinska Institute,
Stockholm, Sweden

1. Perutz, M.F. *Nature* 326, 555–558 (1987).
2. Schrödinger, E. *What is Life?* (Cambridge University Press, 1946).
3. Timoféeff-Ressovsky, N.W., Zimmer, K.G. & Delbrück, M. *Nachrichten auf der Biologie der Gesellschaft der Wissenschaften Göttingen* 1, 189–245 (1935).
4. Kimura, M. *The Neutral Theory of Molecular Evolution* (Cambridge University Press, 1983).

Telling one particle from another

An argument about the proper derivation of the Maxwell–Boltzmann distribution raises questions of the distinguishability of particles with a bearing on whether quantum mechanics is really classical.

WHAT the textbooks say about the distribution of particles (which may be atoms) among the various energy states accessible to them may be over-simple. That is the most obvious conclusion to be drawn from an article by Domenico Costantini of the University of Genoa in *Physics Letters A* for 7 September (123, 433; 1987). The article is a much-delayed riposte to that of J. Tersoff of Bell Laboratories (before it was renamed AT&T Bell Laboratories) and David Bayer of Columbia University (*Phys. Rev. Lett.* 50, 553; 1983).

The underlying issue is whether, and in what circumstances, particles can be regarded as distinguishable from each other. The general opinion, of course, is that classical particles are distinguishable from each other, but that quantum particles are not. There are still many zealous men and women who hope to show that quantum particles are merely classical particles in disguise.

The best place to start is with the textbooks, and with a system consisting of N particles each of which may be in any one of M different states. A particular (microscopic) state of the system may be represented by specifying the state in which each particle is to be found; another is to say how many particles there are in each of the M states.

Plainly the second description is more general than the first. The second may, for example, be specified by a set of numbers n_i which count the numbers of particles in the state whose index is i (running from 1 to M). Obviously this set of occupation numbers must always satisfy the condition that their total is N . The system must also satisfy the condition that its total energy is fixed.

The crucial step in the Boltzmann argument is to calculate the mean state of the system (and also its fluctuations), which entails the calculation of the probability distribution of the statistic n_i . Luckily, as Boltzmann discovered, there are some simplifying assumptions that can plausibly be made. If (as they were in his time) all particles are distinguishable, but if there can be no physical difference between states of the system in which individual particles are interchanged, it is possible to set about the task of calculating the likelihood of particular macroscopic states (specified by sets of occupation numbers n_i).

The problem reduces to that of calculating the number of ways in which a group of N marbles (or objects of any other kind) may be distributed among M containers in such a way that there are n_1 in the first, n_2 in the second, and so on. The standard high-school answer is that the number of ways in which this may be done is $N!/n_1!n_2!\dots$ where the denominator of this fraction is the product of all the factorial functions of the occupancy numbers. Such a configuration is known as a “complexion” in the language of statistical mechanics.

The usual treatment of the problem is to find the state with the greatest likelihood subject to the constraint that the number of particles is given and that the energy is fixed. It is an essential part of the arithmetic that the particles are distinguishable from each other, for otherwise the numbers would not turn out to be the coefficients of the multinomial expansion (the obvious generalization of the binomial expansion) that leads, for example, to Boltzmann’s conclusion that the mean value of the occupancy number is a decreasing exponential function of the inverse temperature.

If it is assumed that particles are indistinguishable from each other, the combinatorial coefficients come out differently and Bose–Einstein statistics are the consequence. The further assumption that two particles cannot be in the same state (after allowing for the possibility of spin) leads to Fermi–Dirac statistics.

That is what we all know. The surprise is that Costantini shows that we may be wrong. He demonstrates that, while the textbook argument and its assumptions may be sufficient, they are by no means necessary. Indeed, he argues, it is also sufficient (in the sense in which mathematicians use the word) to assume that the probability that a particular particle will be found in a particular state is independent of the states in which the other particles have already been assigned to states (subject only to the energy constraint) and that the identity of the particle itself is irrelevant.

Going further, Costantini sets out to replace even these sparse assumptions by what is, in effect, a set of axioms such as Euclid might have devised. Stripped of its subtleties (of which there many), the argument boils down to the assertion that

the probabilities of the occurrence of particular configurations are algebraically symmetric with respect to the interchange of particles — which is not the same thing as saying that the particles are distinguishable.

But if it is not necessary to suppose that particles are distinguishable to arrive at Maxwell–Boltzmann statistics, from what do quantum statistics (Bose–Einstein or Fermi–Dirac) arise? In operational terms, the different possibilities correspond to different ways of choosing the numerical constant in optimization problems subject to a constraint, as in almost any variational problem.

Two features of this argument need comment, of which the most obvious is the contradiction between Costantini and Tersoff & Bayer. Taking a similarly ascetic point of view, the latter set out to show that the assumption in quantum statistical mechanics that particles are indistinguishable can be traded for the assumption that, even with the energy constraint, different configurations (represented by different sets of numbers n_i) will not be equally likely to arise. They say the assumption is plausible, but it is not self-evidently simple.

Two important questions arise, of which the most obvious is that even to those (like most of us) who side with Costantini in believing that quantum mechanics and classical mechanics are different, there must surely be ways in which the question of whether particles are distinguishable or not should be settled by an appeal to the quantum theory of measurement. After sixty years of the Heisenberg Uncertainty Principle, one would expect there to be a way of telling what recipe should be used in what circumstances.

There is also a question of inference to be settled. The Maxwell–Boltzmann distribution is evidently a good representation of reality in the sense that it works. But how rigorously does it have to be tested before it can be taken to be a kind of truth? On the face of things, Costantini’s paper is strong support for the proposition that the distribution is reality, but then there arises the nagging doubt that a description of the real world derived from mere axioms may be a special case of something more general and more simple.

John Maddox

Archaeology

Who were the first Americans?

Jared M. Diamond

ON 12 October each year, Americans celebrate Columbus's 'discovery' of the New World in 1492. In fact, Columbus and subsequent European travellers found the entire western hemisphere populated by Amerindians, whose ancestors are generally agreed to have arrived on foot from Siberia when the low sea levels of the Pleistocene converted the present Bering Straits into a land bridge. Humanity will never again enter a New World with such expanded *Lebensraum*, unless we succeed in reaching another planet. The unresolved debate ranging over that real discovery of America was explored at a recent symposium*.

Artists' renditions of America's discovery generally depict half-naked pre-Neanderthaloid brutes waving crude clubs. But the Bering land bridge was an arctic tundra with few edible plants. Only skilled big-game hunters possessing tailored clothing, warm footwear and methods for winter food storage could have survived there. The Eurasian tundra was first occupied by anatomically modern *Homo sapiens*. Securely dated human artefacts do not appear at the western end of the land bridge (Siberia) until about 25,000 years before present (BP) and at the eastern end (Alaska) until about 12,000 BP. Between Alaska and the contiguous United States lay the barrier of the Canadian ice sheet, which parted to yield a long, narrow, north/south, ice-free corridor around 11,700 BP (S. Porter, Univ. Washington, Seattle). Within a few centuries, undisputed evidence of human presence south of the ice sheet appeared in the form of the distinctive stone weapons called Clovis 'fluted' points, named after the type site in New Mexico where they were defined.

Clovis points mark a culture that flourished briefly but ubiquitously, with little variation, from Alaska throughout all 48 contiguous US states to Mexico (V. Haynes, Univ. Arizona, Tucson). All ¹⁴C-dated Clovis sites in the western United States cluster between 11,500 and 11,000 BP. Sites in the eastern United States are slightly younger, and soon after 11,000 BP evidence of humans appeared throughout South America to the tip of Patagonia. Archaeologists were initially astounded that Clovis hunters and their descendants could sweep 10,000 miles southwards and populate an entire hemisphere within 1,000 years. But for a band of people to increase in number from 100 to a million within a millenium requires a population

growth rate of only 1.4 per cent per year, exceeded in modern times by many expanding populations, such as the Bounty mutineers on Pitcairn island. To sweep 10,000 miles in a millenium requires movements of only 10 miles per year, attested by modern Bantu migrations and surpassed by Eskimo migrations.

Mammoth bones are the most conspicuous prey item at Clovis sites, but other large extinct mammals are also represented, together with small prey such as rabbits, turtles, mice, snakes and lizards. Terminal dates of the Clovis horizon are close to those for New World mammoths, suggesting that Clovis hunters exterminated mammoths (P. Martin, Univ.

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Pomo Indian women at Kashia Rancheria. (Courtesy of Lowie Museum of Anthropology, University of California, Berkeley.)

Arizona, Tucson; L. Agenbroad, Flagstaff) — unless mammoths died out because of climate changes (E. Lundelius, Univ. Texas, Austin) — and then switched to concentrate on bison (D. Stanford, Washington; G. Frison, Laramie).

Were Clovis hunters the first Americans? There have been scores of claims of pre-Clovis sites, most of them grossly flawed by inadequate proof of three essential points: that the site predates Clovis; that some artefacts at the site were made by humans; or that the human artefacts and the dated materials have the same age (R. Morlan, Univ. Ottawa). At present, the best candidate for a pre-Clovis site is Meadowcroft rock shelter in Pennsylvania, meticulously excavated by James Adovasio and colleagues (Pittsburgh Univ.). Meadowcroft has clear stratigraphy, an unquestioned long record of human occupation, and a series of 50 radiocarbon dates that increase in age consistently from the uppermost to the

lowermost strata. The lowest cultural stratum contains distinctive unfluted tools not in the Clovis style and rests on a stratum devoid of human artefacts and dated at 29,000 BP. Carbon-14 samples date that cultural stratum at 13,000 BP or older — significantly earlier than the Clovis horizon found elsewhere (but not at Meadowcroft).

'Clovis-first' archaeologists suspect confounding factors such as contamination of those ¹⁴C samples by carbon in ground water or coal, churning of deposits by burrowing mammals and inconsistent plant associations. Adovasio has counter-arguments to each objection. More ¹⁴C samples have been dated at Meadowcroft than at any other Clovis site. Why do Clovis-first archaeologists question the dates of Meadowcroft's lowest cultural layer, but not of the later cultural layers and earlier non-cultural layer between which the lowest cultural layer lies?

There is a simple answer: Meadowcroft remains unique in its seemingly secure evidence for pre-Clovis Americans. Similar debates in the past about early human occupation of Europe and Australia were settled when proponents satisfied the sceptics' demand for compelling evidence at not just one but at many sites (D. Grayson, Univ. Washington, Seattle). In North America itself, the Clovis horizon is attested by a dozen ¹⁴C-dated sites and by tens of thousands of tools scattered over the whole continent. Pre-Clovis levels at Clovis sites lack evidence of humans. Numerous caves in the western United States suitable for human occupation have yielded abundant dung of extant mammoths and ground sloths radiocarbon-dated to before 12,000 BP, but no evidence of humans. Meadowcroft's pre-Clovis inhabitants, if they existed, could not have arrived by helicopter from Alaska; they had to leave evidence on the intervening terrain. Why has such evidence escaped detection?

Clovis-first archaeologists answer that it does not exist (Martin, P. *Nat. Hist.* No. 10, 10–13; 1987). Pre-Clovis advocates instead say that pre-Clovis settlers could have remained very few in number, or had few and unsophisticated tools, or may not have been big-game hunters. These explanations strain the credulity at least as much as the claims that the meticulous archaeological studies at Meadowcroft are in error. People who were not big-game hunters or lacked sophisticated tools would have frozen or starved to death on the Bering land bridge. There is no precedent for humans reaching rich unpopulated lands full of game and failing to multiply — unless the initial colonist band itself died out, and the existence of Meadowcroft eliminates that argument for pre-Clovis advocates.

Still missing from the pre-Clovis/Clovis-first debate is any effort by either side to

*Americans before Columbus: Ice Age Origins Smithsonian Institution, Washington DC 26 September 1987.

assess statistically the likelihood of detection of pre-Clovis sites besides Meadowcroft, given a reasonable model of population movements from Alaska to Meadowcroft; the abundance of Clovis evidence; and the relative archaeological 'visibilities' of Clovis sites and putative Meadowcroft-like pre-Clovis sites. Such

calculations, along with further scrutiny of Meadowcroft's dating and further efforts to locate other pre-Clovis sites, will eventually provide convincing evidence for who were the first Americans. □

Jared M. Diamond is Professor of Physiology at the University of California Medical School, Los Angeles, California 90024, USA.

Chemical physics

Reactions with directed motion

Richard M. Noyes

THE Belousov-Zhabotinskii (BZ) reaction probably exhibits the richest phenomenology of any known non-living collection of chemicals. Whenever we think we know at least the list of qualitative behaviour patterns in space-time, something new is added to the compendium. Such an addition is the report of Noszticzius *et al.* on page 619 of this issue¹, showing that a train of chemical waves can be made to travel unidirectionally around a ring almost indefinitely somewhat as an induced electrical current can flow around a superconducting coil.

The BZ system involves the oxidation of an organic substrate by acidic bromate, often catalysed by an oxidation-reduction couple, such as Mn(II)-Mn(III) or Ce(III)-Ce(IV), based on a 1-equivalent change (see ref. 2 for a recent review). Over wide ranges of composition, the state of such a system can be characterized as oxidized or reduced, depending on the potential of a platinum electrode or on the colour of a redox indicator. Reduced states have much larger concentrations of bromide ions than do oxidized states.

In some stirred solutions, an oxidized or a reduced state is repeatedly converted suddenly to the other type by 'relaxation oscillations'. In other stirred solutions, a persistent state can be 'excitable';

perturbation beyond a sharply defined threshold will initiate an excursion of several orders of magnitude in the concentration of bromide ion. Temporary and permanent bistability, oscillations induced by silver ions, bursting, amplified fluctuations and deterministic chaos are terms that have evolved to designate some of the more exotic phenomena in open and closed stirred systems.

If a closed system is unstirred, wave phenomena may be generated. Undoubtedly the most dramatic such behaviour is the 'trigger wave', studied in great detail by Winfree³, which consists of a narrow band of oxidized state moving through a medium of marginally stable, but excitable, reduced state. At the leading edge of the band, diffusion depletes the local concentration of bromide and triggers an autocatalytic excursion that temporarily converts the reduced state to an oxidized one. By the time that local region has returned to a reduced state with enhanced bromide concentration, the oxidized region has moved on. A trigger wave moves unidirectionally into a reduced region by destroying bromide ions in front of it and leaving a greater bromide concentration behind. The greater the bromide concentration in front of the wave, the slower it travels: a sufficiently great concentration can stop the wave entirely.

When trigger waves meet, they annihilate each other without diffraction effects. Figure 1 shows two spiral waves twisting in opposite directions and disappearing where they meet. Winfree has generated still more complex three-dimensional patterns and has suggested analogies to arrhythmias in cardiac arrest.

Until now, studies of trigger waves have been conducted in closed systems. Noszticzius *et al.*¹ in this issue report the development of a clever ring system in which reducing chemicals are replenished by diffusion from the centre and oxidizing chemicals from the outside. They also use temporary concentration gradients to block waves travelling in one direction. They thus select waves that are free to circulate around the ring indefinitely in only one direction. Figure 3 (c and d) on

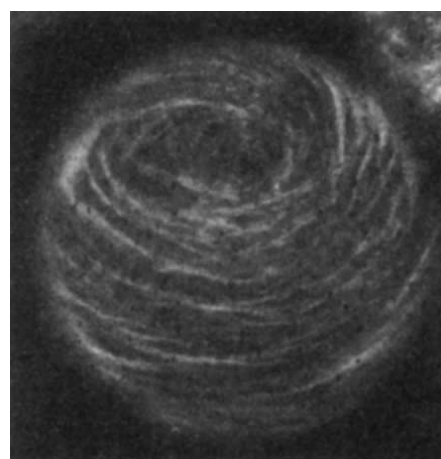


Fig. 2 Microtubules labelled with fluorescent antibodies in a fertilized sea urchin egg. (Courtesy of Patricia Harris, from ref. 4.)

page 620 shows that the waves travel faster near the outside of the ring where the medium is more oxidizing, and that the waves disappear as one moves toward the more reducing medium in the centre. These observations are consistent with those previously reported by Winfree.

Such directed behaviour is reminiscent of chemical streaming found in biological systems. All the wave phenomena observed in the BZ reaction so far, however, are based on the coupling of understood chemical reactions with diffusion of species in less than micromolar concentrations. The BZ system has not yet demonstrated directed motion of macroscopic amounts of material such as are sometimes seen in living cells.

Figure 2 shows a spiral arrangement of microtubules in a fertilized sea urchin egg⁴. These microtubules are apparently responsible for the rotation of the peripheral cytoplasm with relation to the inner region of the cell⁵. Other dramatic examples of movements within cells include the migration of pigment granules in chromatophores of fish scales; actin-based streaming of cytoplasm of certain algal cells; microtubule-dependent movement of chromosomes at mitosis; and transport of vesicles along the length of nerve axons. If bodies big enough to be seen in a microscope move unidirectionally through a viscous fluid, they must be acted on by directed forces of a kind that Newton could have understood. The chemical mechanisms of some of these biological processes are only just beginning to be elucidated. □

1. Noszticzius, Z., Horsthemke, W., McCormick, W.D., Swinney, H.L. & Tam, W.Y. *Nature* **329**, 619-620 (1987).
2. Field, R.J. & Burger, M. (eds) *Oscillations and Travelling Waves in Chemical Systems* (Wiley, New York, 1985).
3. Winfree, A.T. *When Time Breaks Down* (Princeton University Press, 1987).
4. Harris, P., Osborn, M. & Weber, K. *Expl Cell Res.* **126**, 227-236 (1980).
5. Schroeder, T.E. & Battaglia, D.E. *J. Cell Biol.* **100**, 1056-1062 (1985).
6. Winfree, A.T. *Accs chem. Res.* **10**, 215 (1977).

Richard M. Noyes is in the Department of Chemistry, University of Oregon, Eugene, Oregon 97403, USA.



Fig. 1 Rotating spiral trigger waves in a thin layer of BZ solution. Note that where the waves meet, there is no interference pattern. (Courtesy of Arthur T. Winfree, from ref. 6.)

Molecular epidemiology

Reliable exposure assessment

David E. G. Shuker

THE reliable and sensitive detection of characteristic molecular markers for human exposure to chemical carcinogens is no longer a barrier to further study, as several sophisticated analytical techniques, such as tandem mass spectrometry and fluorescence line-width narrowing spectrophotometry, are now available. The evaluation of the risk of developing a tumour following such exposures, however, remains difficult. These two points were the main conclusions of a recent meeting* where cancer epidemiologists and laboratory-based researchers tried to identify areas of common interest.

The term 'molecular epidemiology' is relatively recent (see Perera, F.P. & Weinstein, I.B. *J. chron. Dis.* **35**, 58; 1982), but the area of research has its roots in the pioneering work of the Swedish radiobiologist Lars Ehrenberg. In the late 1950s he, with Ave Gustafsson, drew the government's attention to the potential carcinogenic hazards of human exposure to the radioactive alkylating agents ethylene oxide and ethyleneimine. Ethylene oxide was, and still is, used widely as a sterilant for medical equipment. Ehrenberg and Gustafsson's predictions were based on the chemical and mutagenic properties of these compounds, but the first epidemiological reports, showing increased numbers of leukaemia cases in workers exposed to ethylene oxide, did not appear for 20 years.

DNA damage

In the interim, Ehrenberg and his group had developed an approach to molecular dosimetry based on the adducts formed between chemically reactive species and the erythrocyte protein globin, which can easily be obtained and analysed. Ethylene oxide-globin adducts are formed in a dose-dependent manner and, because of the lifetime of the erythrocyte, cumulative low-level exposure can be detected. More recently, ethylene oxide adducts have been found in globin from smokers at levels related to the number of cigarettes smoked. Ehrenberg and his group have tried to quantify the risk of developing cancer following ethylene oxide exposure by using the rad-equivalence approach. It is possible to calculate the dose of ionizing radiation that would damage DNA to the same extent as a dose of ethylene oxide. As the cancer risks in man of low levels of radiation are well known, this allows an accurate, quantitative assessment of the risks

from ethylene oxide. But it is not known if this approach would work for other chemicals.

The globin-dosimeter approach can also identify hitherto unknown alkylating agents. The alkylated amino-terminal valine residue in the amino-acid sequence of globin can be selectively cleaved by a modified Edman reaction. There are several *N*-alkyl valines in human globin, some related to higher homologues of ethylene oxide and others of unknown origin (S. Osterman-Golkar, Stockholm University). One intriguing aspect of the protein-adduct analyses is that various research groups consistently find low levels of adducts even in nominally unexposed subjects. For ethylene oxide, this 'background' is probably related to ethylene exposure, such as that resulting from inhaling car-exhaust fumes.

A striking example of the power of the protein-adduct approach is related to tobacco smoking. The adduct of the bladder carcinogen 4-amino-biphenyl (4-ABP) and albumin is a sensitive indicator of smoking status, and levels in smokers are much higher than those in non-smokers. All of the many non-smoking human samples that have been analysed contain measurable levels of 4-ABP-albumin, suggesting either that passive exposure to cigarette-smoke carcinogens is widespread, or that there are other sources of 4-ABP (S.R. Tannenbaum, Massachusetts Institute of Technology).

In principle, a more direct approach to detecting exposure carcinogens is to measure concentrations of DNA adducts. For many alkylating carcinogens, characteristic DNA adducts are known to cause mutations (including the kind that convert certain proto-oncogenes into oncogenes). Until very recently, however, DNA adducts could be detected *in vivo* only in experimental animals treated with high doses of carcinogens, and usually required large amounts of DNA (several milligrams) extracted from the target organ.

For monitoring of humans, DNA can be obtained from readily accessible cells, such as peripheral lymphocytes or from placental tissue. The amounts of DNA available from such sources are usually small (10–100 µg), but these quantities are sufficient for highly sensitive techniques such as ³²P-post-labelling, which involves digestion of DNA to 3'-nucleotides which are enzymatically 5'-phosphorylated with high-specific activity ³²P-labelled phosphate. The resulting labelled nucleotide 3', 5'-diphosphates are separated by chromatography. For carcinogens

of high relative molecular mass, the adducted nucleotides have sufficiently different chromatographic properties to allow them to be separated from unmodified nucleotides by thin-layer chromatography or HPLC (K. Randerath, Baylor College, Houston; N. Gorelick, Massachusetts Institute of Technology).

Human exposure to a complex mixture of carcinogens, such as cigarette smoke, results in an array of products that are not present in non-smokers. Unlike many other methods, the ³²P-post-labelling procedure allows the presence of unknown adducts to be detected. An apparent aromatic amine adduct has been detected in tumour tissue of 43 per cent of colorectal cancer patients, at the level of about one adduct in 10⁸ nucleotides (D.H. Phillips, Institute for Cancer Research, London). Workers exposed to styrene had detectable levels of adducts (about 1 in 10⁸ nucleotides) in peripheral lymphocyte DNA, whereas unexposed controls did not (S.M. Rappoport, University of California, Berkeley).

Future prospects

An alternative approach is to determine levels of excised nucleic-acid adducts in urine — in principle, a completely non-invasive method of monitoring. The adducts appear in urine following efficient excision repair of DNA. For example, the main DNA adduct of aflatoxin B₁ (*N*⁷-guanyl)aflatoxin B₁ is excreted via the urine, and has been used to assess exposure to this mycotoxin in China (J. Groopman, Boston University). An important finding from this study is that there seem to be large differences between the way in which different individuals handle the same amount of dietary carcinogens, suggesting that pharmacokinetic factors profoundly influence the degree of DNA damage.

The power of new mass-spectrometry techniques was demonstrated by the example of the detection of several hitherto unrecognized urinary alkylated purines in human urine samples. When small aliquots of a human urine extract are inserted into a tandem mass spectrometer, selection of certain ions from the first mass spectrometer, followed by further fragmentation in the second, demonstrates the presence of *N*⁷-ethyl- and *N*⁷-hydroxy-ethyl-guanine, together with other, as yet unidentified, guanine derivatives (P.B. Farmer, MRC Toxicology Unit).

Looking to the future, a large prospective study of primary hepatocellular carcinoma in Thailand, which will combine molecular and cancer epidemiology, is now under way (X. Bosch, IARC, Lyon). The incidence of this type of cancer among hepatitis-B antigen (HBA)-positive carriers is relatively high, but it seems likely that additional exposure to chemical carcinogens (such as

*Detection methods for DNA-damaging agents in man: applications in cancer epidemiology and prevention, Espoo, Finland, 2–4 September 1987.

aflatoxins) is required. When sufficient cases of hepatocellular carcinoma have appeared in a cohort of HBA-positive carriers, previously collected and stored samples of blood and urine will be analysed for the presence of specific markers of exposure to carcinogens. If exposure to a defined chemical carcinogen (or group of carcinogens) is implicated, the same cohort could be used for intervention studies, whereby removal of the exposure should reduce the risk of hepatocellular carcinoma.

One case where human exposure to defined doses of alkylating carcinogens results in tumours is that of second cancers resulting from chemotherapy for childhood leukaemia. It has been calculated that in Europe and North America there

may be as many as 750 new cases annually of rapidly fatal, iatrogenic leukaemia per annual cohort of Hodgkin's disease patients, who have received treatments with alkylating agents. To eliminate this tragic side-effect of an otherwise efficacious therapy, it is necessary to elucidate the mechanisms of carcinogenic (as opposed to carcinostatic) action of drugs such as cyclophosphamide. The sort of rapidly advancing techniques of DNA-adduct identification and quantification that were described at the meeting, in combination with cancer epidemiology, are needed to address this problem (J. Kaldor, IARC). □

David E. G. Shuker is at the International Agency for Research on Cancer, 69372 Lyon, Cedex 08, France.

Solar System

Organic chemicals in comets

T.R. Geballe

THE generally accepted dirty-snowball theory of comets, formulated by Fred Whipple 37 years ago, predicts that comets are frozen mixtures of dust and various gases and that they formed out of material in the outer reaches of the primordial solar nebula. Observations of cometary material, therefore, can provide information on the original composition of the nebula and on the early evolution of the Solar System. Last year, space probes demonstrated that organic molecules are present in comet Halley^{1,2}, a frequent visitor to the inner Solar System. At least some of these molecules were also detected in Halley in several ground-based infrared spectroscopic measurements³. Now, Allen and Wickramasinghe report in a paper on page 615 of this issue⁴ the ground-based detection of similar organic molecules in the coma of comet Wilson (see figure), which is visiting the inner Solar System for the first time.

Their evidence lies at infrared wavelengths near 3 μm , a spectral region where vibrationally excited C-H bonds radiate strongly. The exact wavelengths of the vibrations depend on the other atoms to which the carbon atoms are attached. In comet Halley, and now in comet Wilson, the observed chemicals emit between about 3.3 and 3.5 μm ; the principal emission feature is approximately 0.1 μm wide and peaks near 3.36 μm . This is different from the profiles of features in this wavelength band, also attributed to organic molecules, seen in various interstellar environments (ultraviolet-excited nebulae, the diffuse interstellar medium and dark clouds).

Unfortunately, the number of hydrocarbons is so overwhelming and many of their spectra are so similar (as well as

complex) that unique identifications are very difficult, especially when observed spectra are limited by low spectral resolution and low signal-to-noise ratios, as they have been in comets. In the case of comet Wilson, detailed reports of higher-resolution spectra will soon be published (A.T. Tokunaga, personal communication), but they are unlikely to provide identifications. It is not even clear whether the molecules

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Comet Wilson, showing the short dust tail pointing towards the south-west and the ion tail ($> 3^\circ$ long) pointing to the south-east.

emitting the 3- μm feature are free or are embedded in solid matter.

Allen and Wickramasinghe suggest⁴ that the spectra of the organic materials in comets Wilson and Halley, and hence the materials themselves, are actually slightly different. They also hypothesize a sequence in the degree of chemical modification by starlight of a basic initial galactic mixture of organic materials, whose resultant spectra extend from that of comet Wilson to those seen in the diffuse interstellar medium, with comet Halley's spectrum somewhere in the middle. Both of these conjectures can be debated, and not only on arguments of signal-to-noise difficulties. It is dangerous to infer the

emission spectrum of a material on the basis of its absorption spectrum, and vice-versa, if the material itself is unknown. But, whether or not Allen and Wickramasinghe's suggestions are correct, perhaps the most interesting and basic result of their work is that the spectra of the two comets closely resemble each other.

When a comet approaches the Sun, the outer surface is vapourized and blown away, exposing fresh material. The diameter of comet Halley, for example, decreases by a few metres per orbital period. The similarity of the 3- μm spectra of comets Wilson and Halley therefore implies that the organic materials ejected from the surfaces of the youngest comets are similar to those that were deep within comet Halley when it formed. There are at least two possible explanations for this. The 3- μm emission feature in comet Wilson might have originated in molecules that were not altered significantly by exposure to cosmic rays or ultraviolet light during the 4.6×10^9 years that they spent on the surface of the comet while it was within the Oort cloud (where all comets formed). Alternatively, the emitting molecules in both comets could have formed only as the comets approached the Sun, after the destruction of similar parent molecules. It is well known that many molecular species observed in comets are the result of dissociation of other species by solar ultraviolet radiation.

A more basic question, still far from being answered completely, is how closely the chemical constituents buried in comets resemble the original molecular cloud material out of which the Solar System formed. It seems likely that the compositions should differ to some extent because, as the outer regions of the primordial solar nebula were heated by the forming Sun and as the comets were heated internally during their formation, the most volatile materials would have been lost from the dust grains out of which the comets were forming⁵. To answer these questions more quantitatively requires detailed laboratory and observational studies and modelling. Obviously, space missions that sample cometary materials on and below the surface, would be the most direct approach, but these are unlikely for the next decade. Nevertheless, with infrared astronomical spectroscopists willing and increasingly able to make ground-based measurements of comets, a steady stream of some of the necessary data should be forthcoming. □

- Combes, M. *et al.* *Nature* **321**, 266-268 (1986).
- Kissel, J. *et al.* *Nature* **321**, 336-337 (1986).
- Baas, F., Geballe, T.R. & Walther, D.M. *Astrophys. J.* **311**, L97-L101 (1986).
- Allen, D.A. & Wickramasinghe, D.T. *Nature* **329**, 615-616 (1987).
- Yamamoto, T. in *Ices in the Solar System* (eds Klinger, J. *et al.*) 205-219 (Reidel, Dordrecht, 1985).

T.R. Geballe is at the United Kingdom Infrared Telescope, 665 Komohana Street, Hilo, Hawaii 96720, USA.

Ecosystems

Nitrogen cycles in perspective

S.P. Long and D.O. Hall

THE findings of perhaps the most detailed and complete whole-ecosystem study so far have recently been reported*. Flows of nitrogen and carbon have been determined for four different arable systems at Kjettslinge farm, near Uppsala, Sweden, over a five-year period. The results have important implications in understanding the relative importance of different soil processes in carbon and nitrogen flow; the significance of different cropping systems to losses of nitrogen; and developments in the study of the ecology of arable lands.

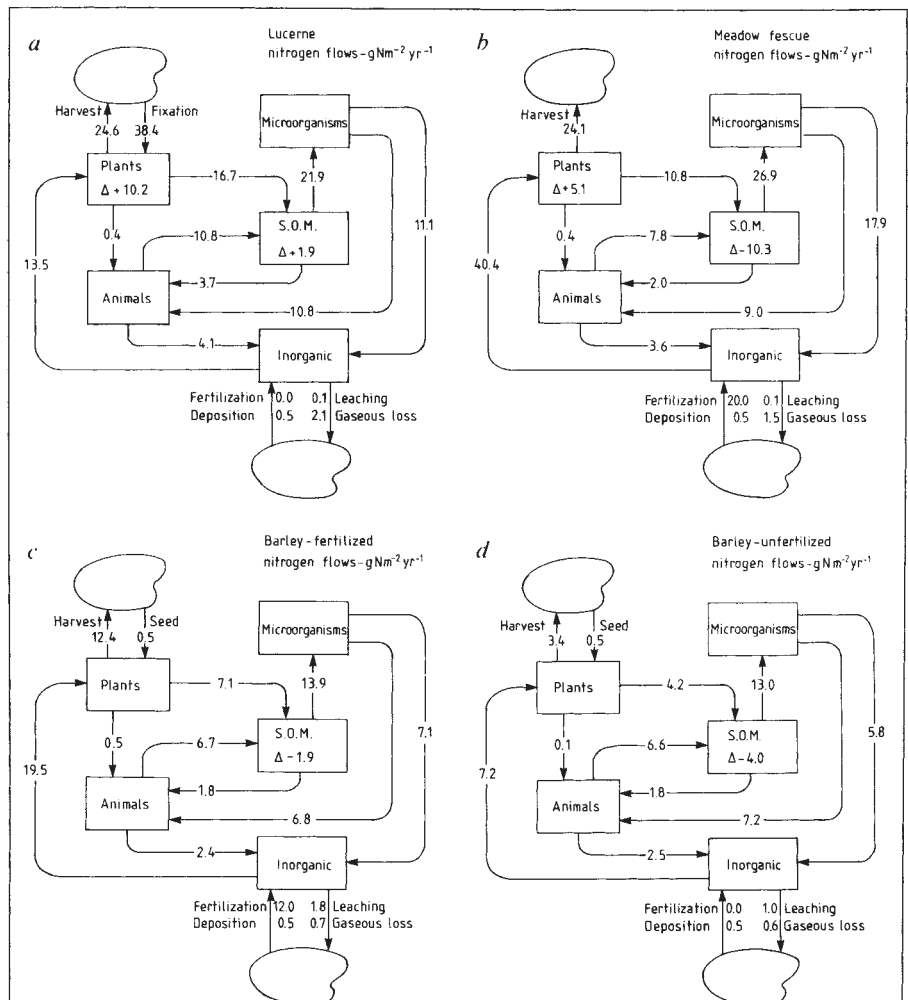
In the ecosystems approach, a community is treated as a functioning entity that can be studied by determining the exchanges of energy and elements between plants, animals, microbes and non-living components that comprise the community. This approach reveals the relative importance of the different components and allows the effects of removing or altering components to be predicted. Although conceptually attractive, the logistics of this approach are formidable, so detailed budgets for any community are rare.

There is an urgent need for an improved understanding of arable lands as ecosystems, especially with respect to the flow of nitrogen and carbon. In the past few decades there has been a continuous increase in the land areas on which cereals are grown and in the grain yields of cereals throughout western Europe, the former mainly at the expense of grassland, and the latter largely resulting from increased levels of inorganic nitrogen fertilization. Inevitably, this situation has led to increases of nitrogen in the environment (J.A. van Veen, ITAL, Wageningen). There is little hard evidence of direct damage to human health resulting from elevated levels of nitrate in diets, but effects on natural ecosystems in some areas have been demonstrated (P. Newbould, HFRO, Edinburgh). Previous studies of nitrogen and carbon flows have been limited largely to measurements of productivity and nitrogen contents of grain and straw at different fertilization rates, which provide insufficient information for analysing ecological relationships. Simultaneous analysis of these elements in microorganisms, soil fauna, plant-root dynamics and root exudates have rarely been made, and mineralization of nitrogen and gaseous losses of nitrogen have not been measured directly (K. Paustian and T. Rosswall, Swedish Univ. Agric. Sci.). The study at Kjettslinge rectifies

these deficiencies and provides a comparison of nitrogen flows both between cereal and herbage crops, and between fertilized and unfertilized crops on the same soils.

The systems examined at Kjettslinge were barley with and without nitrogen fertilizer, a fertilized grass ley, and a lucerne ley (see figure). The results suggest that leaching losses into drainage water are increased more by conversion from ley to cereal than by fertilization, at least in the short term. Losses of nitrogen are greatest from the fertilized cereal system ($1.8 \text{ g m}^{-2} \text{ yr}^{-1}$), equivalent to 1.8

tonnes over 1 km^2 of land and amounting to 15 per cent of the nitrogen added as an inorganic fertilizer (c in the figure). Reduction in fertilizer input would not completely remove this problem, however, as without any added fertilizer the cereal crop still loses $1 \text{ g m}^{-2} \text{ yr}^{-1}$. Losses from both the nitrogen-fixing ley (a in the figure) and the fertilized grass crop (b) are an order of magnitude lower, suggesting that ley crop ecosystems are much more effective at retaining nitrogen even though amounts of nitrogen in all components of these ecosystems are higher than in the cereal systems (Paustian). The cereal systems (c , d) show considerably lower fluxes of nitrogen between soil components, reflecting a lower input of this element into the soil as organic matter and reduced microbial consumption and mineralization in the cereal systems.



Mean annual fluxes in $\text{g nitrogen m}^{-2} \text{ yr}^{-1}$ ($= 10 \text{ kg ha}^{-1} \text{ yr}^{-1}$) for: *a*, lucerne (*Medicago sativa*) ley with symbiotic nitrogen fixation; *b*, meadow fescue grass (*Festuca pratensis*) ley receiving $200 \text{ kg (N) ha}^{-1} \text{ yr}^{-1}$; *c*, spring barley (*Hordeum distichum*) receiving $120 \text{ kg (N) ha}^{-1} \text{ yr}^{-1}$; and *d*, spring barley without nitrogen fertilizer. The leys were maintained for 4 years (1981–84) during a 6-year rotation, with fertilized barley grown in 1980 and 1985. The two barley treatments were maintained as continuous monocultures during the 6 years. Values for lucerne and meadow fescue leys were based on measurements in 1982 and 1983; values for barley were based on measurements from 1981–85. Compartments: 'plants' include all above- and below-ground parts of crops and weeds; 'animals' include protozoa, nematodes, enchytraeids, earthworms, micro- and macroarthropods; S.O.M., non-living soil organic matter; Δ , net annual changes in organic nitrogen.

*Ecology of Arable Lands—The Role of Organisms in Nitrogen Cycling University of Agricultural Sciences, Uppsala, Sweden 9–12 June 1987.

Microbial biomass constitutes about 170 g carbon m⁻², the input of carbon being sufficient for just one new generation of microorganisms per year (J. Schnürer, Swedish Univ. Agric. Sci.). Soil animals constitute about 1–5 g carbon m⁻² — earthworms represent more than half of this biomass. Although they form only a fraction of soil biomass, the animals ingested over a quarter of the carbon input to the soil (B. Söhlenius, Univ. Stockholm) and have an annual turnover of nitrogen similar in magnitude to that of the microorganisms (see figure), emphasizing the important roles of both groups, despite their disparity in quantity. In the grass ley (b) microorganisms and animals mineralize 2.5 times the quantity of nitrogen mineralized in the unfertilized cereal (d), but the proportion directly attributed to animals is more than halved, suggesting that the activity of soil animals assumes relatively greater importance in arable ecosystems of lowered nitrogen status. The high rate of nitrogen fixation in the lucerne (about 38.4 g m⁻² yr⁻¹) and the ability of the system to retain about 95 per cent of this gain illustrates the potential for sustaining crop and soil nitrogen levels with legume–*Rhizobium* associations (a).

The Kjettslinge study provides details of carbon and nitrogen flows for several years at one site. The between-year variation, coupled with the parallel climatic measurements, will provide more clues about the basis of temporal variation. The outstanding question is the extent to which the conclusions of this study are applicable to other parts of Sweden and western Europe, and to other cereal and ley crops. The only comparable study is a Dutch collaborative project in which different fertilizer inputs and crops are being examined, which will be completed in 1991 (van Veen).

Both studies will provide invaluable databases for continued development of models for prediction of the effects of changes in the arable ecosystem and understanding of how fertilization practices can be adjusted to maximize crop uptake and minimize damaging losses to the environment. Validation and improvement of these models needs many comparable studies over a wider range of soils, climates and crops. As the Kjettslinge study represents more than 100 person-years of effort, the expense of similar programmes will inhibit most western European countries, even though it may represent a minute fraction of expenditure on the subsidies that promote and continue to maintain high fertilizer-input cereal production. □

S.P. Long is in the Department of Biology, University of Essex, Colchester CO4 3SQ; D.O. Hall is in the Department of Biology, King's College London (KQC), Campden Hill Road, London W8 7AH, UK.

Molecular evolution

Too much neutral polymorphism

Julee Greenough and Paul H. Harvey

To what extent is molecular evolution simply the result of random mutation and genetic drift, as the neutral theory suggests? This seemingly straightforward question is difficult to answer, because explanations based on natural selection can invariably be constructed for data that conform to the predictions of the neutral model (for example, refs 2, 3). An ingenious test developed by Hudson, Kreitman and Aguadé⁴, however, identifies a useful class of data — that which can be shown to be inconsistent with the neutral hypothesis.

Under the neutral model, the rate of nucleotide substitution in a population is determined by the neutral mutation rate per zygote. Regions of the genome that are constrained by selection have a lower neutral mutation rate (fewer of the mutations that occur are neutral), and are therefore expected to change more slowly over time. This means that populations diverge more slowly at these regions. The theory also predicts that a lowered neutral mutation rate should decrease the level of neutral polymorphism within a population. Combining these two predictions eliminates the need to know the neutral mutation rate: loci that diverge more slowly between populations should be less variable within populations.

It is this prediction that forms the basis of the new statistical test of the neutral theory⁴. Natural selection will cause deviations from this prediction, because it alters the level of neutral polymorphism in a population but has no effect on the rate of neutral change (see ref. 5). The maintenance of a selected polymorphism in the population enhances neutral polymorphism⁶ because the fixation of linked neutral mutations is hindered, whereas the fixation of a selectively favoured allele annihilates the neutral variation associated with other alleles.

Hudson *et al.*⁴ ask whether the levels of neutral polymorphism at the alcohol dehydrogenase (*Adh*) locus and adjacent 5' flanking region in the fruitfly *Drosophila melanogaster* are consistent under the neutral model with the rates of change in these regions between *D. melanogaster* and its sibling species *D. sechellia*. Inter-specific comparison reveals comparable levels of divergence at the two loci (the flanking sequence differs at 5.2 per cent of the 4,052 nucleotide sites, whereas the *Adh* locus differs at 5.6 per cent of 324 silent sites). This implies that the mutation rates in the two regions are comparable, so we would expect to find similar levels of neutral polymorphism at each site.

To determine the extent of polymorphism, Hudson *et al.* use data from the 'four-cutter' technique of DNA analysis⁷ to examine 81 independently derived, isochromosomal lines of *D. melanogaster*. This technique uses a battery of enzymes acting at specific, four-nucleotide-long sequences to break up genomic DNA. The DNA fragments are then denatured, separated by electrophoresis, and probed with a radioactively labelled DNA segment containing the *Adh* locus. Used with restriction maps based on complete DNA sequences of the region, this technique can detect virtually all insertion/deletion variation and about 25 per cent of all single nucleotide substitutions. The results show that the level of silent polymorphism at the *Adh* locus is about four times greater than in the flanking region. (Of a possible 79 silent sites in the *Adh* locus at which nucleotide substitution could have been detected, 10.1 per cent were variable, compared with 2.2 per cent of 414 possible sites in the flanking region.)

Hudson *et al.* develop a goodness-of-fit statistic to determine whether these data represent a significant departure from the neutral model. They find that the neutral hypothesis is rejected at a probability level of 0.016, even though the assumptions made in developing this test (for example, that no recombination occurs within loci and that the loci are unlinked) are favourable to the neutral hypothesis. A *post hoc* explanation for the results under the neutral model requires that the neutral mutation rate in the *Adh* locus of *D. melanogaster* has increased relative to that of *D. sechellia* since the divergence of the species. Hudson *et al.*, however, find that their test is quite insensitive to changes in this mutation rate. In fact, the probability of a false rejection of the neutral hypothesis does not reach 50 per cent until the mutation rate for one region in one species is more than six times that for the same region in the other species.

Thus, the patterns of polymorphism and divergence observed at the *Adh* locus and adjacent 5' flanking region in *Drosophila* are inconsistent with the neutral theory. What can we make of this deviation from neutral expectations? Both the flanking region and the silent sites of the *Adh* locus are thought to be relatively unconstrained, and were therefore expected to diverge at the same rate and show similar amounts of silent polymorphism. Because the data agree with the first expectation but not the second, the rejection of the neutral model seems

to result from the observed patterns of polymorphism. As the polymorphism of the 3' flanking region, which contains another coding sequence, is only one-third that in the *Adh* locus, and as most of the polymorphism at the *Adh* locus is in one exon, the problem seems to be an excess of polymorphism in the *Adh* locus.

The excess silent variation is consistent with the presence of a selected polymorphism in the *Adh* region. In fact, two electrophoretically distinct alleles are known for the *Adh* locus, and there is some evidence that the polymorphism may be selectively maintained⁸. Further, of the three exons comprising the coding sequence of *Adh*, the one containing the amino-acid change that distinguishes the two alleles has the highest level of silent polymorphism. The picture is not quite perfect, however: it seems there is still an excess of silent polymorphism within one type of allele⁹.

Hudson *et al.* have developed a useful test for detecting the action of natural selection on molecular evolution. It will be interesting to see how the neutral model fares when the test is applied to

other loci. For example, Hill and Hastie recently presented data¹⁰ on a class of enzyme-inhibiting proteins in which the rate of nucleotide substitution causing amino-acid replacements in the reactive centre is greater than the rate of synonymous substitutions in the rest of the sequence. If the rapid divergence of the reactive centre is caused by selection for parasite resistance, as the authors¹⁰ suggest, then a comparison of the type performed by Hudson *et al.*⁴ should reveal reduced neutral polymorphism in the reactive centre. □

1. Kimura, M. *The Neutral Theory of Molecular Evolution* (Cambridge University Press, 1983).
2. Skibinski, D.O.F. & Ward, R.D. *Nature* **298**, 490-492 (1982).
3. Gillespie, J.H. *Oxf. Surv. Evol. Biol.* **4**, 10-37 (1987).
4. Hudson, R.R., Kreitman, M. & Aguadé, M. *Genetics* **116**, 153-159 (1987).
5. Kreitman, M. *Oxf. Surv. Evol. Biol.* **4**, 38-60 (1987).
6. Strobeck, C. *Genetics* **103**, 545-555 (1983).
7. Kreitman, M. & Aguadé, M. *Proc. natn. Acad. Sci. U.S.A.* **83**, 3562-3566 (1986).
8. Oakeshott, J.G. *et al.* *Evolution* **36**, 86-96 (1982).
9. Kreitman, M. E. & Aguadé, M. *Genetics* **114**, 93-110 (1986).
10. Hill, R.E. & Hastie, N.D. *Nature* **326**, 96-99 (1987).

Julie Greenough and Paul H. Harvey are in the Department of Zoology, University of Oxford, Oxford OX1 3PS, UK.

Ocean Drilling Program

New studies of the Indian Ocean

*Leg 115 shipboard scientific party**

LEG 115 is the first of a nine-leg programme of exploration of the Indian Ocean, which hitherto has not been well studied. Our seven-week voyage had two main aims: to verify the proposed hotspot track linking the Deccan flood basalts in western India with young-ocean-island volcanism at Reunion; and to investigate the Neogene history of carbonate production and dissolution in the tropical waters of the Indian Ocean.

One of the most striking geological features of the Indian Ocean basin is the large number of elevated ridges and plateaux. It has been variously proposed that these are fragments of microcontinents, drowned island arcs or the volcanic products of transform faulting or hotspot activity. The Ninetyeast Ridge and the Mascarene Plateau-Chagos-Maldives-

Laccadive Island Ridge are two such volcanic lineaments, which are parallel and could record the northwards motion of India over stationary hotspots at Kerguelen and Reunion, respectively (see map). If so, the massive accumulations of Deccan flood basalts, which erupted about 65 million years (Myr) ago, at the time of the Cretaceous/Tertiary boundary, may be the first manifestation of the Reunion hotspot.

We drilled at four sites (706, 707, 713 and 715; see map) to reach basement rocks along the proposed Reunion hotspot track. Except at its southern end, this ridge is completely covered by thick (200-500 m) deposits of carbonate ooze, and by bank and reef deposits which, on top of the plateau, are of the order of 1 km thick. Biostratigraphic ages determined from the oldest sediments recovered at three sites (35 Myr at 706, 47 Myr at 713 and 55-60 Myr at 715) show that the volcanic rocks conform to the hotspot model of increasing age of volcanism northwards towards India. We drilled at site 707 to determine the southeastern extent of Precambrian granites forming the Seychelles islands, but we found Palaeocene basalts that probably erupted as part of the Deccan flood-basalt event and rifted away from India with the Seychelles block in the early Palaeocene, about 64 Myr ago.

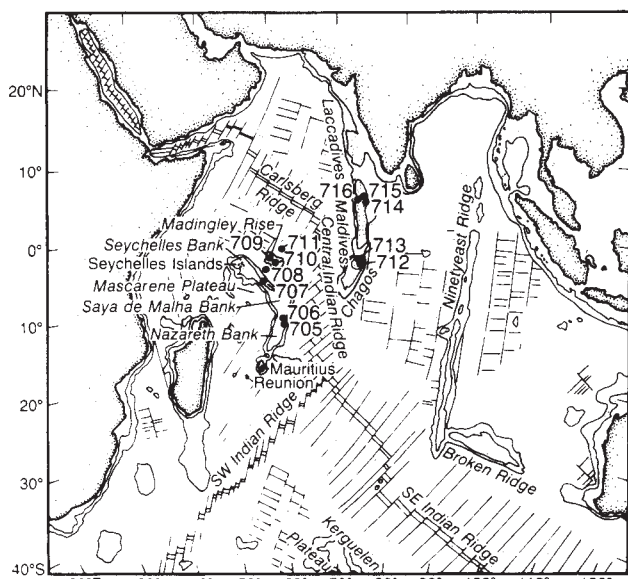
The basalts at these four sites appear to

come from mixtures of magmas from both hotspot and spreading-ridge melting regions. This is not unexpected, because of the close proximity of the hotspot and the central Indian Ridge during formation of the Chagos Bank (site 713) and northern Mascarene Plateau (site 706). In fact, the hotspot started beneath the Indian plate, was crossed by the spreading ridge about 35 Myr ago, and has since been beneath the African plate. A similar history has seen the Kerguelen hotspot over-run by the south-east Indian Ridge and separated from its volcanic trail, the Ninetyeast Ridge. Site 715 basalts, recovered at the northern end of the lineament in the Maldives archipelago, are geochemically similar to basalts from Reunion and thus were probably formed in an intraplate setting.

A further result from the basement drilling along this lineament is our identification of the true polar wander during the Cenozoic. Hotspots remain relatively stationary with respect to one another, and their volcanic trails provide a reference frame for plate motions (see the recent News and Views article by Peter Olson, *Nature* **327**, 559; 1987). Unless the hotspot frame has moved with respect to the geomagnetic pole (assumed to coincide with the Earth's rotational axis), it follows that each volcano formed over a given hotspot should have the same palaeomagnetic latitude. Any departure from a constant palaeolatitude can be considered true polar wander, or motion of the entire Earth's mantle with respect to its spin axis. The Leg 115 basement sites give palaeolatitudes of about 28° S, or 7° south of the present-day Reunion hotspot, indicating the amount of true polar wander of site 706 over the past 35 Myr.

Calcium carbonate produced by unicellular microorganisms in the ocean's surface layer eventually ends up as calcareous oozes, which cover about half (140 million km²) the ocean floor. Two factors, more than anything, determine the distribution of such oozes: the rate of production of sediment-forming microorganisms (mainly planktonic foraminifers and coccolithophores); and their dissolution at depth. Although it is analytically easy to map the stratigraphical variation of carbonate content in deep-sea sediments, the factors that control carbonate accumulation are not well understood. On Leg 115 we studied the interaction between the flux of carbonate production and the dissolution of this material as a function of water depth, giving particular attention to the changing climate and ocean circulation during post-Eocene times (the past 36 Myr). We also studied aragonite (carbonate)-bearing oozes from the Maldives archipelago, to investigate links between the climate and the concentration of aragonite in the sediment, and their underlying causes.

*Co-chief Scientists: Jan Backman (Univ. Stockholm) and Robert Duncan (Oregon State Univ.). ODP Staff Scientist: Andrew Macdonald (ODP, Texas A&M Univ.). Also Paul Baker (Duke Univ.), Alistair Baxter (City of London Polytechnic), Anne Boersma (Stony Point, New York), James Cullen (Salem State College), Andre Droxler (Rice Univ.), Martin Fisk (Oregon State Univ.), John Greenough (St Mary's Univ.), Robert Hargraves (Princeton Univ.), Peter Hempel (Univ. Kiel), Mike Hobart (Lamont-Doherty Geological Observatory), Michael Hurley (Institute of Oceanographic Sciences, Godalming), David Johnson (Woods Hole Oceanographic Inst.), Naja Mikkelsen (Geological Survey of Denmark), Hisatake Okada (Yamagata Univ.), Larry Petersen (Univ. Miami), Domenico Rio (Univ. Parma), Simon Robinson (Univ. Liverpool), David Schneider (Lamont-Doherty), Peter Swart (Univ. Miami), Yoshiyuki Tatsumi (Kyoto Univ.), Didier Vandamme (Inst. de Physique du Globe), Gustav Vilks (Atlantic Geosci. Centre), Edith Vincent (Univ. Pierre et Marie Curie).



Location of Leg 115 sites in the central Indian Ocean.

The history of carbonate budget changes can be revealed by the mass accumulation rates of the biogenic carbonate, the non-carbonate fraction and the bulk sediment. These rates were obtained from a series of sites along a bathymetric transect ranging in water depth from 1,541 m (site 707) to 4,424 m (site 711) on the northern Mascarene Plateau and the Madingley Rise.

All sites, except the deepest (site 711), exhibit similar patterns of bulk accumulation rates. Deposition rates during the Oligocene (36–24 Myr ago) were relatively slow (about $0.5 \text{ mg cm}^{-2} \text{ yr}^{-1}$), whereas the early and middle Miocene (24–12 Myr ago) was a period of pronounced carbonate dissolution, associated in part with slump and turbidite deposition and strongly reduced sediment-accumulation rates. Accumulation rates rose sharply (to about $1 \text{ mg cm}^{-2} \text{ yr}^{-1}$) from about 10 Myr ago to the present, except at site 711. These rates presumably reflect the establishment of modern productivity and circulation patterns in the western tropical Indian Ocean.

The sedimentary column recovered at site 707 shows signs of extensive sediment winnowing, reflecting the vigorous flow of intermediate-depth waters across the northern Mascarene Plateau. Judging from the sediment accumulation rates alone, such currents were particularly important from the latest Oligocene to the middle Miocene. A stepwise reduction in accumulation rates, starting during the late Pliocene (5 Myr), is associated with the onset of large-scale glaciation of the Northern Hemisphere, indicating a direct link between the climatic deterioration and increased flow rates of intermediate-depth water masses.

We chose site 707, which is 1,541 m below sea level, to show carbonate-accumulation patterns that have been virtually unbiased by the effects of carbonate

dissolution. So we were surprised to find low-amplitude (90–95 per cent by mass), long-period ($\sim 0.4\text{--}0.5$ Myr) oscillations in carbonate content over the past 30 Myr. The carbonate content also varies cyclically in the deeper sites (708–711), but with larger amplitudes. The deepest site (711) was below the depth of the carbonate-saturation horizon from about 20 to 7 Myr. Only small amounts of nannofossil carbonate were preserved after the latest Miocene (about 8 Myr), indicating a slight deepening of the carbonate-saturation horizon. On the other

hand, the earliest Miocene and Oligocene show nannofossil carbonate contents averaging about 80 per cent by mass.

Two sites were drilled in the Maldives archipelago to recover aragonite-bearing oozes deposited near shallow carbonate banks. The deeper of these sites (714, at 2,038 m) yielded only a short sequence (19.5 m) of recent oozes (0–0.5 Myr old), because of a hiatus from 0.5–8 Myr. The sequence from before the late Miocene (8–26 Myr), however, shows good preservation of carbonate and siliceous microfossil groups, stratigraphical continuity and also uncomplicated and relatively high sedimentation rates. At the shallower of the two sites (716, 544 m deep), about 260 m of aragonite-bearing sediment was penetrated. This thick continuous sequence was apparently deposited in the past 5.5 Myr, at a rate that exceeded 100 m Myr^{-1} during the first million years. In this section, lower pelagic oozes have become chalk. The oozes contain aragonite throughout and in the deeper parts big lumps of celestite (SrSO_4), thereby providing an excellent opportunity to improve our understanding of post-deposition changes of deep-sea sediments. □

Biochemistry

Polymorphic regulation of membrane lipid composition

Ben de Kruijff

How are the lipids in a cell membrane organized? The answer to this question will allow the nature of the proteins that sense the state of the membrane lipids, and the mechanism by which they direct the enzymes regulating their composition, to be elucidated. Recent work by Goldfine and colleagues¹ provides a basis for these questions to be addressed.

The bilayer structure originally proposed by Gortel and Grendel in 1925 is still an extremely attractive way to explain many properties and functions of membranes. But two properties of membrane lipids have for a long time been difficult to explain in terms of this model: first, membranes are typically composed of more than 100 different lipid species, although a single species alone could form a bilayer; and second, how can the many classes of polymorphic membrane lipids that preferentially adopt a non-bilayer configuration be accommodated into the model?

In recent years, the basis has been laid for understanding the enormous diversity of membrane lipids. Low-abundance membrane lipids have essential specialized functions, such as permitting communication between cells and allowing them to respond to their environment. This, together with the fact that many membrane proteins contain specific covalently bonded (phospho)-lipids, and that virtually all the main classes of

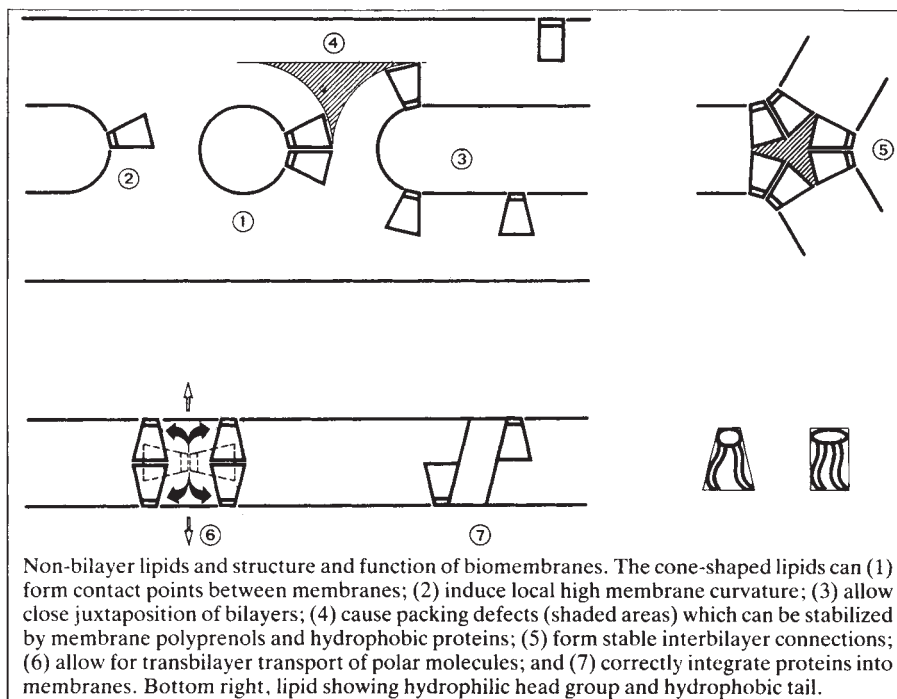
membrane lipid have unique biophysical properties, can help us appreciate the necessity of such diversity. At first, 'fluidity' was thought to be a key element in regulating lipid composition, but now it appears that cells control their overall lipid composition by lipid polymorphism.

Polymorphism is the unique ability of hydrated lipids to aggregate into different types of structures. The pioneering X-ray diffraction studies² of Luzatti and colleagues, recently extended by nuclear magnetic resonance³ and electron-microscope⁴ methods using well-defined lipid systems, established that every biological membrane contains large amounts of non-bilayer lipids next to bilayer-forming lipids. On isolation and hydration under physiological conditions, these lipids adopt non-bilayer structures, of which the inverted type II hexagonal phase H_{II} , a hexagonal assembly of tubes in which the lipid head groups surround a narrow aqueous channel, is the most common. The abundant monogalactosyldiglyceride from plant chloroplasts is an example of this type of lipid^{5,6}. It is difficult to reconcile membrane function with extensive areas of hexagonal H_{II} phase, so it is not surprising that this phase has been observed only in biomembranes under pathological conditions. But in model systems under conditions intermediate between those resulting in bilayer and H_{II} phases, so non-

bilayer structures, such as lipidic particles⁷ and cubic phases⁷, have been seen that are more compatible with membrane structure and function.

The realization that all biological membranes undergo essential processes during which the lipid bilayer must be

membrane lipids such as lysophospholipids and gangliosides prefer to be organized in micelles or other type I (convex water-lipid interphase) structures because these molecules have their head group at the wide end of the cone. The formation of bilayers in mixtures of oppositely shaped



(transiently) disrupted, for example membrane-fusion endocytosis and exocytosis, coupled with the fact that inverted non-bilayer structures resemble the morphology and dynamic behaviour of biomembrane systems such as the tight junction⁸, chloroplast⁹, prolamellar body¹⁰ and liver microsomes¹¹, imply that non-bilayer lipids are involved in these processes and formation of such structures in membranes. The modulation of bilayer/non-bilayer transitions by several parameters (including ions, drugs, antibiotics, peptides and (membrane) proteins¹²; the requirement of non-bilayer lipids for the activity of several membrane-bound enzymes^{13, 14}; and the correlation between the presence of non-bilayer lipids and membrane functions such as fusion, transbilayer transport of divalent cations, lipids and proteins¹⁵) substantiated this idea and shed new light on the function of polyprenols in membranes¹⁵ and the mechanism of action of lipid-seeking and structure-modulating drugs and toxins.

The molecular basis of polymorphism is surprisingly simple. The shape-structure concept¹ relates the dynamic shape of a lipid molecule to the macroscopic structure of the lipid aggregate. Bilayer-prefering lipids are roughly cylindrical whereas inverted type II (concave water-lipid interphase) structure-prefering lipids are cone-shaped with the head group at the smaller end of the cone (see figure). Detergents and low-abundance

non-bilayer lipids is a convincing demonstration of the validity of the shape-structure concept of polymorphism¹⁶.

This concept can now be extended to lipid-protein interactions since the discovery that the channel-forming polypeptide gramicidin modulates lipid structure. This peptide has a cone shape because of four bulky tryptophan residues at the carboxy terminus, and this can account for H_{II} phase formation with bilayer lipids and bilayer formation with type I lipids¹⁷. Within a bilayer organization, non-bilayer lipids can have important functions (see figure). By virtue of their low head-group hydration they allow the juxtaposition of opposite bilayers; a situation commonly encountered in energy-transducing membrane systems such as chloroplasts, mitochondria and rod outer segments which are particularly rich in non-bilayer lipids. By virtue of their shape, non-bilayer lipids can also determine local membrane curvature or permit the correct integration of bulky intrinsic membrane proteins. It is not surprising that most of the known membrane-spanning alpha-helical segments of complex membrane proteins are tilted with respect to the normal plane of the bilayer, which is likely to lead to packing defects with solely bilayer-forming lipids. And surely the success of asolectin in its ability to reconstitute functional, complex proteins results from the very high non-bilayer lipid content of this lipid mixture.

The various roles of non-bilayer lipids in membranes suggest the bilayer/non-bilayer lipid ratio needs to be closely regulated *in vivo*. Studies on microorganisms now convincingly demonstrate such regulation. *Acholeplasma laidlawii* has been shown¹⁸ by Wieslander, Lindblom and collaborators to regulate strictly the diglycerol/diglyceride/monoglycerol/diglyceride ratio in its membrane in response to changes in environment. These glycolipids are the dominant lipid components of the membrane, the di- and monoglycerol/diglyceride being a bilayer and non-bilayer lipid, respectively. When growth conditions are such that the adoption of non-bilayer structures becomes more favourable, less of the non-bilayer lipid is synthesized, and vice versa.

The new work of Goldfine, Shipley and colleagues¹, on the chemically totally different lipids of the *Clostridium butyricum* cell membrane, shows that in the presence of biotin, membrane lipids became highly enriched in the oleic acid added to the medium. This enrichment results in a marked increase in the amount of glycerol acetal of plasmenylethanolamine, a lipid unique to this group of organisms, at the expense of phosphatidylethanolamine (PE) and the parent plasmalogen plasmenylethanolamine (PlaE). Analysis of the biophysical properties of the hydrated isolated lipids reveals that the increased tendency of the oleic acid-enriched PE and PlaE to adopt inverted non-bilayer phases is efficiently compensated for by the increased amount of the bilayer-prefering glycerol acetal analogue. □

- Goldfine, H., Johnston, N.C., Mattai, J. & Shipley, G.G. *Biochemistry* **26**, 2814-2822 (1987).
- Luzzati, V., Gulik-Krzywicki, T. & Tardieu, A. *Nature* **218**, 1031-1034 (1968).
- Cullis, P.R. & de Kruijff, B. *Biochim. biophys. Acta* **559**, 399-420 (1979).
- Verkleij, A.J., Mombers, C., Leunissen-Bijvelt, J. & Ververgaert, P.H.J.Th. *Nature* **279**, 162-163 (1979).
- Shipley, G.G. in *Biological Membranes Vol 2* (eds Chapman, D. & Wallach, D.F.H.) 1-89 (Academic, London, 1973).
- Sen, A., Williams, W.P. & Quinn, P.J. *Biochim. biophys. Acta* **663**, 380-389 (1981).
- Rifors, L., Erikson, P.-O., Arvidsson, G. & Lindblom, G. *Biochemistry* **25**, 7702-7711 (1986).
- Kachar, B. & Reese, T.S. *Nature* **296**, 464-466 (1982).
- Jacob, J.S. & Miller, K.R. *J. Cell Biol.* **103**, 1337-1347 (1986).
- Simpson, D.J. *Res. Commun. Carlsberg* **43**, 145-170 (1978).
- Van Duijn, G., Luiken, J., Verkleij, A.J. & de Kruijff, B. *Biochim. biophys. Acta* **863**, 193-204 (1986).
- Tilcock, C. *Chem. Phys. Lipids* **40**, 109-125 (1986).
- Jensen, J.W. & Schutzbach, J.S. *Biochemistry* **23**, 1115-1119 (1984).
- Navarro, J., Toivo-Kinnunen, M. & Racker, E. *Biochemistry* **23**, 130-136 (1984).
- de Kruijff, B. et al. in *The Enzymes of Biological Membranes* (2nd edn) Vol 1, (ed. Martonosi, A.N.) 131-204 (Plenum, New York, 1985).
- Rand, R.P., Pangborn, W.A., Purdas, A.D. & Tinker, D.O. *Can. J. Biochem.* **53**, 189-195 (1975).
- Killian, J.A. & de Kruijff, B. *Chem. Phys. Lipids* **40**, 259-284 (1986).
- Wieslander, A., Rifors, L., Johansson, L.B.-A. & Lindblom, G. *Biochemistry* **20**, 730-735 (1981).

Ben de Kruijff is Professor of Biochemistry at the Institute of Molecular Biology and Medical Biotechnology, University of Utrecht, Padualaan 8, 3584 CH Utrecht, The Netherlands.

Vascular system of the giraffe

SIR—Both the discussions of the vascular system of the giraffe you have published recently mention venous valves, but both overlook the special 'seried' valves found at the entry of major tributaries into the axillary and brachial veins, though not into the jugular. Fewer occur in the closely-related okapi and some in the jugular and femoral veins of the bactrian camel¹.

These valves have a cusp proximal to the orifice of entry and another distal to it. Both are concave proximally. A number of slight differences occur in the arrangement of the cusps. Using fresh material from the camel the valve proved to be inefficient when pressure in main vein and tributary were very low, but when pressure in the main vein was raised progressively to as high as 200 mm of mercury regurgitation into the tributary was prevented. This pressure was higher than suggested for within the feet of the giraffe and surely for anywhere in the camel. The original workers had no access to fresh, unopened veins from a giraffe, but presumably their function in that animal would be the same and they would obviously be a great aid to the muscular pump. Why such valves should occur in the neck of the camel is hard to explain. The arterial structures in the neck of the camel² and both of the *Giraffidae*³ differ substantially.

R.E. REWELL

*The Old House,
Vowchurch,
Herefordshire HR2 0RB, UK*

1. Amoroso, E.C., Edholm, O.G. & Rewell, R.E. *Proc. zool. Soc.* **117**, 435 (1947).
2. Ask-Upmark, E. *Acta psychiat. Suppl.* **6**, (1935).
3. Lawrence, W.E. & Rewell, R.E. *Proc. zool. Soc.* **118**, 202 (1948).

Small-particle physics and interstellar diamonds

SIR—Lewis *et al.*¹ have demonstrated the presence of extremely small (50 Å) diamonds in unshocked chondrite meteorites and have convincingly argued that these diamonds were formed in a circumstellar environment as metastable condensates. No consideration has yet been given to the possibility that diamond may be the stable condensate in this size regime and that thermodynamic predictions based on the bulk properties of minerals may not be applicable to nanometre-size particulates. Unfortunately, efforts to correct for the effects of the relatively large surface/volume ratio of particles in this regime are hampered by the lack of reliable data on the surface free energy of appropriate minerals.

Rietmeijer *et al.*² have argued that in the mineral system enstatite–tridymite–for-

terite, for particle sizes ≤ 100 Å in diameter, the mineral mixture tridymite plus forsterite is more stable than enstatite alone, despite the fact that large enstatite grains are more stable than the mixture of large tridymite plus large forsterite grains. The reversal in the relative stability of the assemblage occurs because enstatite grains have a larger surface free energy than either forsterite or tridymite grains. I propose that a similar stability reversal could occur with carbon; for example, surface effects could make diamond the stable mineral phase for very small particles. Unfortunately, this hypothesis is difficult to test as values for the surface energy of both diamond and graphite are extremely unreliable. Bulk graphite is ~ 700 cal mol⁻¹ ($\sim 3 \times 10^{10}$ erg mol⁻¹) more stable than diamond³. The molar volume³ of graphite is 5.298 cm³ whereas that of diamond is 3.417 cm³. A reasonable estimate⁴ of the average surface free energy of a graphite sphere would be between 2,000 and 4,000 erg cm⁻²; diamond surface energies are in the range 3,700–9,800 erg cm⁻². Given the large uncertainty in the surface free energies for diamond and graphite it might be instructive to perform an alternative calculation to see if the hypothesis of a stability reversal appears reasonable for very small carbon particles.

If we assume a value for the surface free energy of graphite and also assume that particles smaller than 50 Å are more stable as diamond than as graphite, we can calculate the value of the surface free energy of diamond for which this set of assumptions would be correct. At equilibrium one mole of 50-Å diameter graphite grains would coexist with one mole of 50-Å diamond particles. But as the molar volume of graphite is larger than that of diamond this yields a surface area of 6.4×10^7 cm² for one mole of graphite grains while the surface area of a mole of 50-Å diamonds is 4.1×10^7 cm². We have therefore

$$(\text{area} \times \text{surface energy})_{\text{dia}} + \Delta G_{\text{dia}} = (\text{area} \times \text{surface energy})_{\text{gra}} \quad (1)$$

As ΔG_{dia} is $\sim 3 \times 10^{10}$ erg, if we assume that the surface energy of graphite is $\sim 4,000$ erg cm⁻², then the diamond surface energy for which equation (1) is correct is $\sim 5,500$ erg cm⁻². If the surface energy of graphite is only $\sim 2,000$ erg cm⁻² then the corresponding diamond surface energy is 2,400 erg cm⁻². In other words, because of the larger number of less dense graphite particles needed to contain one mole of carbon, diamond's surface free energy does not actually need to be smaller than that of graphite for diamond to be more stable and it can even be significantly larger. The consequences and implications of this hypothesis will be explored in another paper⁴.

Finally, it appears obvious that reliable

measurements or calculations of the surface free energies of solid particles are necessary if we ever hope to predict correctly the behaviour of very small particle systems. It should also be obvious that conventional 'thermodynamic wisdom' does not necessarily apply to nanometre-scale grains.

I wish to thank J. Hecht, G. Huss and B. Donn for several useful discussions.

JOSEPH A. NUTH III

*Laboratory for Extraterrestrial Physics,
NASA/Goddard Space Flight Center,
Greenbelt, Maryland 20771, USA*

1. Lewis, R.S., Ming, T., Wacker, J.F., Anders, E. & Steel, E. *Nature* **326**, 160–162 (1987).
2. Rietmeijer, F.J.M., Nuth, J.A. & Mackinnon, I.D.R. *Icarus* **66**, 211–222 (1986).
3. Robie, R.A. & Waldbaum, D.A. *US Geol. Surv. Bull.* **1259** (GPO, Washington DC, 1968).
4. Nuth, J.A. *Astrophys. Space Sci.* (in the press).

Porosity of nuclear fuels

SIR—In interpreting Nichols' most interesting paper¹ on porosity in nuclear fuels, Robert Cahn² has slipped into the trap of considering mechanisms in isolation. He rightly comments on the importance of grain boundary porosity formation as a contributor to fission product gas release from nuclear fuels and the micrograph he shows, produced by Hyam and his co-workers in the UK Atomic Energy Authority Laboratories at Windscale, is one of several which formed the basis of the model³ still in use to predict such release. It is not correct, however, to deduce that fission product gas release in the columnar region is small. The high temperatures and thermal gradients that enable such grains to develop lead also to very high levels of gas release, probably by the bubble sweeping mechanism to which Nichols himself refers.

R. HARGREAVES

*UKAEA, Windscale Laboratories,
Seascale, Cumbria CA20 1PF, UK*

1. Nichols, F.A. *J. nucl. Mater.* **148**, 92–98 (1987).
2. Cahn, R.W. *Nature* **328**, 295–296 (1987).
3. Hargreaves, R. & Collins, D.A. *J. Br. nucl. Energy Soc.* **15**, 311–318 (1976).

How many Chernobyl fatalities?

SIR—Not to be stubborn, but for the sake of scientific truth, I would like to respond to J.H. Fremlin (*Nature* **327**, 376; 1987). The supralinear shape of the dose-response curve both for atomic worker's cancer and for cell depression in mice is not a hypothesis but an observed fact. It correlates additional effects with additional exposures, with the obvious conclusion that the background radiation will have corresponding effects, which appears to be born out by recent careful observations. Thus Fremlin's hypothesis that background radiation somehow immunizes people against further radiation is not born out by the facts.

The "more correct control group" to which I referred as being used in the recent investigations of Hiroshima and Nagasaki cancers is the corresponding national average of Japan, which is more reliable than the 0–9 rad group; assuming the latter to be free of radiation effects (which has been falsified by later experience) obviously induces a circular reasoning leading to an underestimation of the true risk. The evidence obtainable from US vital statistics showing correlations between immunological competence and fallout radiation burden contradicts Fremlin's assertion that radiation can stimulate the immune system.

JENS SCHEER

Physics Department,
University of Bremen,
2800 Bremen 33, FRG

FREMLIN REPLIES—Scheer is wrong in stating that the national average cancer rate in Japan is more reliable as a control than that of the 0–9 rad group from the bombed cities. Big towns like Hiroshima and Nagasaki may well have cancer rates differing from the Japanese average. In Britain for example, Edinburgh has a cancer rate 7% higher than that of Scotland as a whole and 24% higher than the unpolluted Western Isles, even if we leave out lung cancers as being largely the result of cigarette smoking.

The best control for Hiroshima and Nagasaki consists of those inhabitants of the two cities who received zero dose, having been too far away, underground or in the part of Nagasaki shielded by hills, at the time of bombing. They will have had the same average diet and environment as those irradiated, which will not have been the case for most areas of Japan.

In the accompanying table the numbers of cancer deaths (including leukaemia) observed following the bombing, for both sexes in both cities, as a function of the radiation dose received (from Tables II, III and IV of H. Kato and N.J. Schull, *Radiat. Res.* **90**, 395–432; 1982) are compared with the number of cancers predicted. The latter were found by multiplying the zero-dose number of deaths by the ratio P_1/P_0 where P_1 is the number of person-years in the irradiated group concerned and P_0 the number of person-years in the zero-dose group as shown by Kato and Schull in appendix Table II (the figures for both sexes in both cities have been added together).

It is of course uncertain that the age distributions and other characteristics will have been the same for all groups, but this is unlikely to be as important as the statistical uncertainties arising from the relatively small number of excess deaths in any one group. The recorded numbers of deaths in all groups are of course subject to the usual uncertainties familiar in counting statistics, the standard deviation in a

Cancer deaths, including leukaemia, 1950–78, of 82,000 survivors within 10 km of the hypocentres of the Hiroshima and Nagasaki bombs

Dose range (rad)	0	1–9	10–49	50–99	100–199	200–299	300–399	400+
Number of person-years (thousands)	764.4	561.2	361.8	102.8	76.6	34.4	15.7	21.1
Total cancer deaths in study group	1,816 ±43	1,248 ±35	924 ±30	270 ±16	222 ±15	124 ±11	55 ±7	97 ±10
Predicted for same number of person-years for zero dose	1,816 ±43	1,333 ±31	860 ±21	244 ±6	182 ±4	82 ±2	37 ±1	50 ±1
Observed minus predicted		–85 ±47	64 ±37	26 ±17	40 ±15	42 ±11	18 ±7	47 ±10
Observed minus predicted above and below 100 rad			5 ±62			147 ±22		

count of N being $\sqrt{N}$.

The percentage deviations for the predicted deaths are necessarily the same as the percentage deviation of the zero-dose figure, because the number of person-years for each group is accurately known, and will contribute negligibly to errors.

The results are surprising. Simple addition shows that the total number of all cancers up to 1978 attributable to the bombing (observed minus predicted deaths) was 152 ± 66 , but simple addition is inappropriate; nearly all of the error arises from the results of the doses below 100 rad while nearly all of the positive results are due to larger doses.

Starting with the 0–9 rad group, there is a 10% probability that the negative result is a pure chance deviation. Adding the result from the next column we find that there is still a net negative cancer increase of -21 deaths below 50 rad. The probable error is larger than 21, but this result is nevertheless quite important. It means that after 29 years of study of the effects of radiation there is no evidence whatsoever that there is *not* a threshold in the region of 50 rad below which the risk of additional cancer is zero. Indeed, a higher threshold still is not precluded; the total excess of cancer deaths below 100 rad (1 gray) is 5 ± 62 .

Above the 100 rad level there is no question that there is a serious deleterious effect; the 40 apparent excess of deaths for doses between 100 and 199 rad is over 2.5 standard deviations above the predicted 15, giving a confidence limit above 99%, and the probability that the 147 excess deaths calculated for doses above 100 rad have happened by chance is less than one in a hundred million.

Apart from the Japanese results, in none of the low-level studies cited by Scheer was any attempt made to investigate the smoking habits or early environments of victims or controls. Looking for cancer effects from low-level radiation without such investigations is a bit like listening for mosquitoes at a disco.

The increase of lifetime risk of cancer in moving from the Western Isles to Edinburgh is greater than the risk due to 400+ rad at Hiroshima. I suggest that Scheer

compares the age-standardized cancer rate in Bremen with that in any rural area on the Atlantic coast of Europe and uses his considerable persuasive skill to reduce the forms of pollution that really matter.

I wish to make it explicitly clear that these results from Hiroshima and Nagasaki do not prove that there are no harmful effects of even the lowest level of radiation dose. They show only that the bomb results cannot be used to prove that there are harmful effects below 100 rad. This is particularly striking as one would expect that any dose of radiation given in a fraction of a second is more likely to cause observable damage than the same dose spread over months or years. Indeed, I know of no case of any injurious experience, from swallowing 100 mg of strychnine or arsenic to losing 10 litres of blood or falling downstairs, which is less dangerous when all done in a few seconds than when spread over months.

Again, I do not want to imply that any conclusions that are reached from the Japanese results, which are now believed to be almost entirely produced by gamma rays, necessarily apply to the effects of internal irradiation by alpha particles arising from radon or plutonium. And nor do they affect in any way the catastrophic nature of nuclear weapons, the long-term effects of which on health have always seemed to me more worrying than dangerous when compared to their blast and heating effects.

Further cancer deaths are to be expected among both the irradiated and the unirradiated survivors, but there is no reason why the proportions should change greatly. In general, cancers appear more rapidly after larger doses, so that the high-dose figures would be expected to rise proportionately less than the zero dose figures. Provisional figures for deaths up to 1982, for which I thank Dr Schull, have added many thousands to the study group and have increased the proportion of cancer deaths between 10 and 99 rad, but do not change the conclusions.

J.H. FREMLIN

46 Vernon Road,
Edgbaston,
Birmingham B16 9SH, UK

Reflections of wartime

Willem Hackmann

Radar Days. By E.G. Bowen. *Adam Hilger, Bristol/Taylor & Francis, Philadelphia:* 1987. Pp.231. £12.50, \$28.

Radar in World War II. By Henry E. Guerlac. *Tomash/American Institute of Physics:* 1987. Two volumes, pp.1,171. \$110. \$88 for members of ten societies associated with the American Institute of Physics.

PROFESSIONAL historians are becoming increasingly interested in the history (social, political and technical) of military science, a subject that has fascinated the lay reader for many years. Of particular interest is radar, which was largely responsible for the defeat of the German Air Force and the U-boat in the Second World War. Summing up the war at sea, for example, Admiral Doenitz wrote that the U-boat had been deprived of its essential feature — the element of surprise — by means of radar, and that the scientists who had created radar had been called the saviours of their country.

These two books are the latest on the origins and wartime development of radar. Dr Bowen, "the father of airborne radar", has written a personal account of how the first airborne radars were built and how they were introduced into the Royal Air Force. His book has all the advantages and drawbacks of what professional historians refer to grandly as "oral histories".

Professor Guerlac, on the other hand, was a professional historian of science. His massive (and rather indigestible) two-volume history was written as a fully documented, classified report of the wartime radar programme of Division 14 of the National Defense Research Committee (NDRC). It is largely the history of this division's main laboratory, the Radiation Laboratory at the Massachusetts Institute of Technology, though it contains much more besides. It does not deal with radar countermeasures, as these were the responsibility of Division 15 of the NDRC, nor with the research at America's principal electronic laboratories, chief among them the Bell Telephone Laboratories. Even so, the ambitious plan of covering the entire spectrum of the activities of Division 14 has resulted, inevitably, in a patchy and sometimes rather superficial treatment.

Guerlac began in early 1943 with the help of a single assistant, increased to half a dozen by the time his task was completed in 1946. He was recruited to write an official history, and the incentive for commissioning it was to justify the large amount of money spent on radar in case there should be a congressional investigation — the kind of reason that would appeal to bureaucrats and that is not

unknown to 'official historians' on both sides of the Atlantic.

One of the most far-sighted decisions taken by the Office of Scientific Research and Development, the umbrella organization of the NDRC, was to have the wartime developments of the different divisions written up. Those of the Radiation Laboratory were described in a series of 28 technical monographs published by McGraw-Hill. Guerlac's history was rejected for this series as it was not considered technical enough. A condensed version was next offered to Little Brown & Company, who dealt with wartime memoirs, but for a variety of reasons it was not published. The unedited manuscripts have been among the most frequently cited documents in histories of radar and allied subjects since the war.

The present volumes are an edited merger, approved by the author, of both the 'long' and the 'short' versions of his manuscript. It is a pity that Guerlac died before he could write the Introduction because he had always expressed reservations about this work, which he regarded not so much as a history, but the basis for a future history. This hesitation on his part highlights a fundamental controversy among historians concerning modern histories and the 'historical perspective'.

A commonly held opinion that such a perspective is enhanced with time is being undermined by much recent work — as one classical historian has aptly remarked, what we gain when writing about earlier periods is merely confidence in generalizations which we would never dare to make if we had access to the real wealth of contemporary evidence.

Guerlac's book is a mine of what was, at the time of writing, almost contemporary information. It is written in the style of a technical report, though for ease of access the text is produced in sections, further broken up by numerous chapters and bold sub-headings. Volume I is concerned largely with the technical history of radar and covers its early history in the United States and Great Britain, the establishment of the NDRC radar programme, the technical developments at the Radiation Laboratory (both of the individual radar systems, of which more than 100 were developed, and of the main components, such as the various types of magnetron), and ends with the administrative history of Division 14. Volume II is the operational history of radar as a combat weapon. German radar is dealt with in a slight manner in one of the appendices.

The book suffers from its long gestation. References stop in 1945, there is no assessment based on modern research and certain subjects are now dealt with better in more recent histories. On the other hand, Guerlac's work has been of inestimable value in the past, and this monumental study still contains a great deal of value and interest.

Bowen's book, too, is not a straightforward history, but what it lacks in documentary evidence is made up for by the vividness of the author's recollections of past events. The account is written from



How many holes? Lee DuBridge, then Director of the Radiation Laboratory, and Isidor Rabi discuss the resonant magnetron with E.G. Bowen at the Radiation Laboratory in 1940.

the perspective of the individuals who worked at the laboratory bench, and is peppered with anecdotes. For instance, the power of the first air-warning antenna, at Orfordness, was increased to such a level that the arcing to the supports could be heard over a mile away on a quiet night. The cure was effected by means of a couple of copper balls from a lavatory cistern. This solution characterizes much of the early research and development of the British radar team (as of the interwar British military R & D in general); unlike the Americans, they paid little attention to preparing designs ready for industrial mass production.

One of the most fascinating aspects of Bowen's book is his description of the Tizard Mission which initiated Anglo-American military scientific cooperation in late 1940, and introduced to the Americans the resonant cavity magnetron invented by Boot and Randall at Birmingham University. The author was one of the members of this vital mission, and has an extraordinary story to tell.

The magnetron taken to Washington was number 12 of the first production batch. Unknown to Bowen, the specimen which he had taken with him to the States was a prototype in which the original six

cavities had been increased to eight. This caused great puzzlement to the Americans, for he described and showed them the blueprints of the six-cavity device and yet X-rays of their specimen showed two extra cavities. As it worked well the Americans decided to copy this specimen, and this is the reason why all early British magnetrons had six holes and the American ones had eight. It was fortunate that this untried prototype worked so well, for number eleven of this first batch, which had seven holes, did not work at all.

It is perhaps inevitable in personal reminiscences that a certain amount of stereotyping takes place. Tizard emerges as an almost heroic figure, and the author is obviously less keen on Lindemann (Lord Cherwell), A.P. Rowe, the wartime superintendent of the radar establishment, and the management generally.

Both Guerlac's and Bowen's accounts are worthwhile additions to the growing literature on radar history. But I suspect that Bowen's slim volume will give the most enjoyment. □

Willem Hackmann is Assistant Curator at the Museum of the History of Science, Old Ashmolean Building, Broad Street, Oxford OX1 3AZ, UK.

All in the family

P.H. Williams

Microbial Genetics Applied to Biotechnology: Principles and Techniques of Gene Transfer and Manipulation. By Venetia A. Saunders and Jon R. Saunders. *Croom Helm, London/Macmillan, New York*: 1987. Pp.422. Hbk £30, \$40; pbk £14.95.

IN the beginning microbial genetics was exciting, elegant, esoteric. We learned of the nature of genetic variation, the structure of genes, the intricacies of gene regulation, and so much more — and it all fitted into one book, William Hayes's brilliantly written *The Genetics of Bacteria and their Viruses*. Later, microbial genetics took second place in the scientific and public imagination to its precocious offspring genetic engineering, also exciting, also elegant, but theoretically universally applicable, the panacea for all biological problems. Its proponents tended to forget its roots in real genetics, and its texts resembled nothing so much as recipe books. But the prodigal son has returned; certainly recombinant DNA technology is powerful, but alone it cannot always provide all the answers, and indeed it is often impotent in the absence of other approaches.

This, I believe, is the Saunders' message; their excellent book outlines both "old fashioned" genetics and "new

fangled" cloning as experimental sciences equally relevant and mutually supportive in their application to modern biotechnology. Early chapters, on techniques described in principle rather than in experimental detail, provide an invaluable overview of current methodological trends, and although extensive prior knowledge on the part of the reader is assumed, references to reviews and key articles are always at hand for the benefit of the newcomer. Later chapters, on various applications of microbial genetics in industry, medical and veterinary practice, agriculture and environmental sciences, are less convincingly written in that they are of necessity somewhat speculative or prospective. Nevertheless, many outstanding practical examples are presented in some detail, interspersed with occasional prophecies of future developments.

Thus the book provides scientists, technologists and students of applied biology alike a balanced, authoritative view of the far-reaching effects of microbial genetics, ancient and modern, on a range of commercial enterprises. My only disappointment was in the poor standard of figures (the typed annotation was sometimes messy and difficult to decipher) and in the irritation of their often being a turn of the page away from the text that describes them. This useful little book deserves, I think, better service from its illustrations. □

P.H. Williams is a Lecturer in the Department of Genetics, University of Leicester, Leicester LE1 7RH, UK.

The pimply face of AI

Igor Aleksander & Helen Morton

Encyclopedia of Artificial Intelligence, Vols 1 and 2. Editor-in-chief Stuart C. Shapiro. *Wiley*: 1987. Pp.1,219. \$175, £166.

MANY years ago, some research laboratories in computing used the words 'experimental computing' to describe what is now known as artificial intelligence (AI). John McCarthy of Stanford University is said to have been responsible for the more spectacular phrase which, for the past 30 years or so, has been the bannerhead of a subject that has struggled from controversy to controversy. Is it really about making super-smart machines? Does it provide a theoretical basis for psychology? Does it explode some philosophical myths about the intellectual supremacy of the human species? And would any of these debates have arisen had the field stayed with its original title?

The advent of the *Encyclopedia of Artificial Intelligence* should mark some sort of maturity, some sort of stability for this topic. After all, an encyclopaedia is supposed to be a complete (albeit concise) source of educative information on its subject. Certainly the Japanese Fifth Generation computing project, and the fact that much AI in the United States is now funded by the US Department of Defense, suggests that the subject is being taken sufficiently seriously for industrialists to seek cost-effective applications and for the military to hope for greater autonomy for their weapons' systems. Does the *Encyclopedia* provide evidence of stability? Even a cursory glance suggests precisely the opposite.

Herb Simon, a doyen of the field, defines AI in his guest editorial as "Pro-

PDP in paperback

One of the publishing events of 1986 was the appearance of *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* by David E. Rumelhart, James L. McClelland and the PDP Research Group. MIT Press have now issued the two-volume work in paperback. The books are respectively entitled *Foundations and Psychological and Biological Models*, and were reviewed in *Nature* 327, 469-470 (1987). Price in paperback is \$25, £17.95 for the set.

The same authors have written a companion volume, *Explorations in Parallel Distributed Processing: A Handbook of Models, Programs and Exercises* describing how to use their programs and write new ones. The book has accompanying software, and will be published at the beginning of next year. Provisional price is \$27.50, £17.25.

gramming computers in ways that, if observed in human beings, would be regarded as intelligence". This clashes resoundingly with Elaine Rich's definition within the body of the book, under the heading "Artificial Intelligence": "AI is the study of ways in which computers can be made to perform cognitive tasks at which, at present, people are better". The one speaks of human performance as an accomplished achievement within AI, while the other sees it as a target. The one suggests that AI technology is related to human mechanisms, while the other sees the similarity as being in the task rather than the method.

It was therefore courageous of Stuart Shapiro to have rounded up some of the busiest people in computer science and faced them with the problem of explaining, in a non-partisan way, the key components of AI. The result is somewhat patchy, reaching heights of excellence in some of the contributions and the limits of superficiality in others. For example, having heard that PROLOG is a good language for AI work, a browser will discover only six lines on the subject. These increase the reader's curiosity by confirming that the language is important, having been adopted as the standard for the Japanese Fifth Generation programme. The entry then gives no further details. This can be contrasted with the 16 or so pages that are devoted to LISP: older but American-invented, where PROLOG is distinctly European in origin.

Altogether, this is virtually an all-American affair: of the 200 contributors only nine come from outside the United States, none from Japan. Two of the non-

Americans (colleagues from Britain) write extensively about logic programming, putting right some of the neglect of PROLOG. (This points to a general problem: you have to know where to look to find the appropriate reference.) The entry on logic programming is also the only place in the *Encyclopedia* where the substantial work done in the British Alvey project is mentioned. We labour this point not out of patriotic fervour, but from a belief that a one-nation outlook cannot be complete in a subject that many of the leading industrial nations of the world have singled out for special support. The sins of omission are numerous, discussion of AI architectures being particularly unbalanced. Sweden's successful pyramidal vision systems get no mention and neither do various array and adaptive systems developed in Britain.

Despite the chauvinism, there is much in the *Encyclopedia* that is excellent. We put ourselves in the position of a researcher at the beginning of a trawl of the AI literature. As an example of a somewhat neglected area of research, we chose "Language Acquisition" and found an excellent survey. The choice of references is appropriate, not padded and thoughtfully divided into the specific and the general. Particularly impressive is the reference to connectionist models of language acquisition which are drawn from a revival of research in neural networks that is growing with astonishing rapidity in the United States.

Doing things the neural way has become fashionable only in the past few years, so the *Encyclopedia* is up to date, at least in this area. Looking up "Con-

nectionism" in its own right provides an informative article on this topic describing some of the hopes that AI researchers have for attacking experiential as opposed to rule-dominated knowledge. But it fails to refer to the origins of the revival: the work of John Hopfield at the University of California at San Diego. It also fails to point out that interest in this field has only been *rediscovered* recently, but may be traced right back to the work of McCulloch and Pitts in 1943.

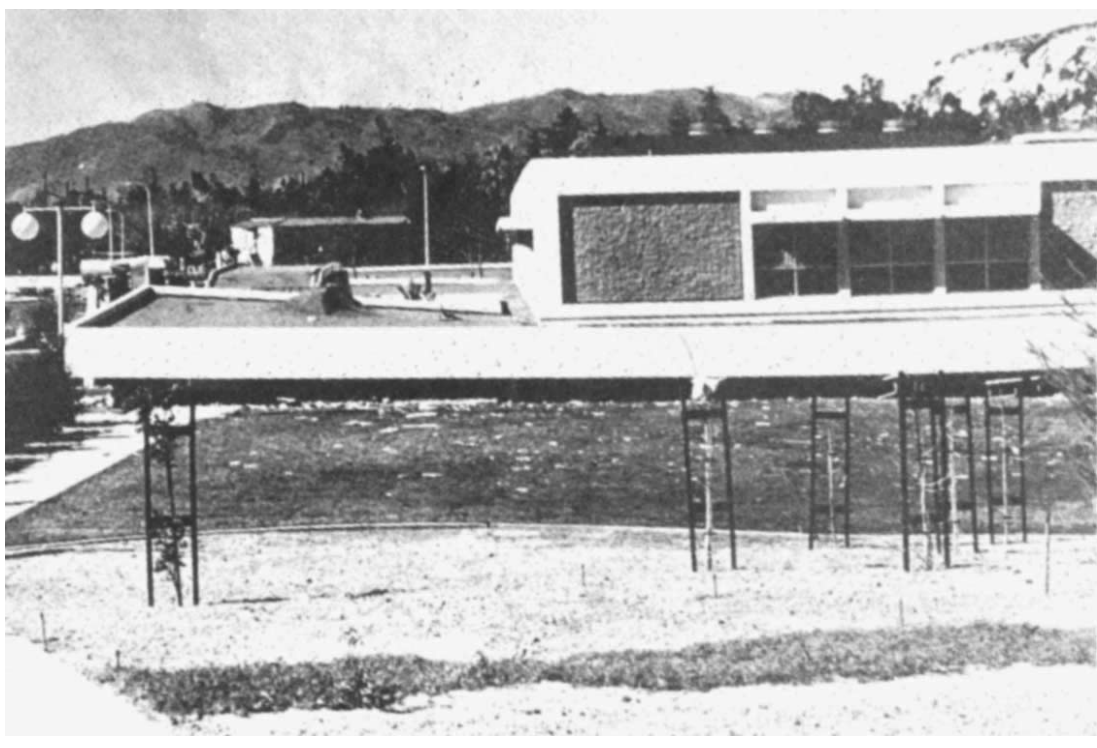
Unfortunately, not all entries are equally up to date. A look at the now well-trodden area of expert systems reveals a description that refers mainly to systems built in laboratories and that could well have been written some years ago. It fails to discuss the rapid recent development of empty systems, or 'shells', which those who are building their own expert systems find so helpful.

Despite the weaknesses and the poor cross-referencing, this is an essential piece of equipment for any AI laboratory. In particular, the budding PhD student will find most useful the ready-made literature searches and the potted statements about important directions for development. But were we to buy these rather expensive books, we would like the assurance of some kind of an updating service. In a rapidly changing field such as this, the gaps and changes of emphasis must grow exponentially with time. □

Igor Aleksander is Kobler Professor of Information Technology, Department of Computing, Imperial College, University of London, London SW7 2BZ, UK. Helen Morton is a Lecturer in the Department of Human Sciences, Brunel University, Uxbridge, Middlesex UB8 3PH, UK.

Down to earth — the Psychiatric Clinic at the Community Hospital, Olive View, California, after the 1971 San Fernando earthquake. The building, made of reinforced concrete, was originally two storeys high, but the ground floor collapsed completely and the first floor here rests upon the debris. No one was seriously hurt in the incident.

The picture is taken from the timely revised edition of Bruce A. Bolt's popular book *Earthquakes*, just published by W. H. Freeman. Price is hbk \$25.95, £22.50; pbk \$13.95, £12.50.



Bruce A. Bolt

Disease overview

P.L. Pearson

Duchenne Muscular Dystrophy. By Alan E.H. Emery. *Oxford University Press*. 1987. Pp.315. £35, \$55.

IN HIS foreword to this latest addition to the series *Oxford Monographs in Medical Genetics*, Sir John Walton expresses the opinion that single-disease monographs can never be expected to be best-sellers. I sincerely hope that this book proves to be an exception to the rule, and that it will become compulsory reading for all associated with neuromuscular disorders.

In an easily read style, Emery takes the reader through many aspects of muscular dystrophy ranging from its possible inclusion on murals of ancient Egyptian tombs to the recent contributions of recombinant DNA technology. It is possible that experts in any one field will feel Emery has given insufficient attention to his or her speciality to do it justice. However, as Emery himself states in the introduction, it is not his intention to cater solely for particular interests and the book passes rapidly from topic to topic, whetting readers' appetites without boring them, and providing an adequate bibliography.

Following a historical introduction to the disease, the author considers the clinical features and makes extensive use of his own observations, occasionally backed up by series published by other authors. Although at times the number of observations used to demonstrate a particular point is limited, this is not a major criticism because the homogeneity of a single series outweighs the numerical advantage of observations pooled from different sources. It would, however, have been more helpful if similar data had been expressed in a standard format, for example, the data on age of learning to walk expressed in the same format as that used for other age-related phenomena.

Notwithstanding such small presentation hiccups and an occasional typographical error (the most amusing being "grog-nosis"), the book moves smoothly from one topic to another. The text is well documented with interesting figures, and one marvels in particular at the incredible muscular enlargement and definition illustrated in a negro boy with Duchenne muscular dystrophy. The section on confirmation of the diagnosis contains pertinent data on serum levels of enzymes and clear illustrations of the muscular pathology encountered, not only in Duchenne muscular dystrophy but associated disorders also. Herein lies one of the strengths of this book, in that the reader is given enough information to understand both the clinical and genetic grounds for distinguishing Duchenne muscular dystrophy

from other neuromuscular disorders.

Not unexpectedly, Emery covers aspects of the inheritance of the disease, mutation frequency, carrier detection and prenatal diagnosis clearly and succinctly. For those who still do not understand the rationale of combining risks from various sources to estimate carrier risk, this book is a must. The use of DNA markers is considered in just sufficient detail to make the technique clear to non-geneticists without getting them bogged down in minutiae.

The broad basis is kept up to the end. The final sections on genetic counselling and patient management include humane and compassionate discussions of the rights of the parents in counselling decisions and the psychological problems

encountered by the patients themselves.

One feels that in some respects the book has been published just a little bit too early. Rapid molecular developments in characterizing the gene are now changing our conception of what causes the high-mutation frequency and giving insights into the role of the abnormal gene product to account for the differences between Becker and Duchenne muscular dystrophy. However, much of the clinical information will remain relevant and I hope that the author will be encouraged to up-date the molecular sections in the not-too-distant future. □

P.L. Pearson is a Professor in the Department of Human Genetics, Sylvius Laboratories, University of Leiden, Wassenaarseweg 72, 2333AL Leiden, The Netherlands.

Poste standards

Jack A. Lucy

Cell Fusion. Edited by Arthur E. Sowers. *Plenum*: 1987. Pp.540. \$75 (North America), \$90 (elsewhere).

CELL fusion is a feature of the preparation of monoclonal antibodies and, in this rather restricted guise, it is familiar enough. However, for many years the wider aspects of cell fusion and membrane fusion that underlie so many different biological phenomena have been pursued by relatively few investigators. This situation appears to be changing, judging at least by the accelerating rate at which books on the subject are appearing. Six years passed before the publication in 1978 of *Membrane Fusion*, edited by Poste and Nicolson, was followed by a Ciba Symposium volume entitled *Cell Fusion*. Only three years later we have the present book with an identical title, while the proceedings of an international meeting on "Molecular Mechanisms of Membrane Fusion", held in Buffalo in April, are currently in press.

The latest volume is divided into four parts of roughly equal importance: "Fusion in Cell Membranes"; "Fusion in Model Membranes"; "New Membrane Fusion Methods"; and "Applications of Membrane Fusion". The book is an interesting and useful reference source, particularly with regard to some of the recent advances in these fields. It has been published to a high standard (although not entirely free from typographical errors), the micrographs are well reproduced and each chapter has its own bibliography in which the titles of papers are included.

The editor notes in his preface that most current work on membrane fusion is within the context of intact or modified cells, and that the book emphasizes the plasma membrane. He also comments that the material covered in the section on plasma

membranes was selected to emphasize depth rather than breadth. Unfortunately for anyone using the book as an introduction to the field, this selectivity has apparently been responsible for the exclusion of fertilization and the associated acrosome and cortical granule reactions. Fertilization is not even mentioned in the index. The formation of multinucleate cells in inflammatory reactions by the fusion of macrophages (covered by Poste and Nicolson) is also not considered, and membrane fusion in endocytosis is given little attention.

Comparisons with Poste and Nicolson's book are rather invited because the preface refers back to it and remarks that there is an obvious need for an up-to-date monograph that integrates material on new fusion methods and their applications, reviews established material, and rationalizes and integrates the old and the new. In my view, however, the present volume does not approach the bench-mark set by Poste and Nicolson in either depth or breadth. Although it contains eleven more chapters, it is about 40 per cent shorter and several of the contributions are straightforward accounts of new work which would be much more appropriately published in a regular research journal.

The new book is also much less catholic in its representation of international research. Only two of the 26 chapters are from laboratories outside the United States, compared with half of those in Poste and Nicolson's volume. This is particularly noticeable in relation to electrofusion. Many significant papers have been published on the fundamental importance and potential applications of electrofusion by laboratories in Germany and France, but there are no contributions from them among the five chapters on this topic. □

Jack A. Lucy is Professor of Biochemistry at the Royal Free Hospital School of Medicine, University of London, Rowland Hill Street, London NW3 2PF, UK.

Where science has gone wrong

T. Theoharis and M. Psimopoulos

The current predicament of British science is but one consequence of a deep and widespread malaise. In response, scientists must reassert the pre-eminence of the concepts of objectivity and truth.

In 1968 Sir Peter Medawar wrote:

Ask a scientist what he conceives the scientific method to be, and he will adopt an expression that is at once solemn and shifty-eyed: solemn because he feels he ought to declare an opinion; shifty-eyed because he is wondering how to conceal the fact that he has no opinion to declare¹.

The witty style of this passage might amuse the reader. But the humorous tone only serves to conceal the gravity of the matter.

Medawar's remarks were made in the, financially speaking, halcyon days of the 1960s. The British science budget was then increasing by over 10 per cent every five years in real terms². The growth continued in the 1970s, though at a lower rate. But a drastic reversal occurred in the 1980s. During the first five years of this decade, public expenditure on science in Britain declined by over 10 per cent in real terms, and all the signs indicate that the cutbacks will continue. But British scientists cannot complain that they had not been warned. In 1971 Shirley Williams, then a Member of Parliament and later (1976–1979) Secretary of State for Education and Science, spelled out an unmistakable warning:

For the scientists, the party is over. . . . Until fairly recently no one has asked any awkward questions . . . Yet there is a growing suspicion about scientists and their discoveries It is within this drastically altered climate that the dramatic decline in expenditure on scientific research in Britain is taking place³.

Public spending on science has declined in other countries too. But a combination of reasons peculiar to Britain has made the situation perhaps the worst in any advanced industrial state. It is not our intention to discuss here all these reasons. Rather our objective is to identify and endeavour to combat what we consider to be the most fundamental, and yet the least recognized, cause of the present predicament of science, not only in Britain but throughout the world.

In 1986 British scientists reacted to the cutbacks with the launching of a campaign called Save British Science (SBS). Its stated objectives were:

. . . to communicate to the public, Parliament and Government a proper appreciation of the economic and cultural benefits of scientific and technological research and development, and of the consequent importance to the nation of

adequate funding of research by Government and industry.

These are also the declared aims^{3,4} of the Royal Society (RS). As its President Sir George Porter says, the RS "is effectively the national academy of science of the UK and it must therefore assume some special responsibility for science in this country"⁵.

Interestingly, British philosophers were quick to emulate the scientists in launch-

"Having lost their monopoly in the production of knowledge, scientists have also lost their privileged status in society. Thus the rewards to the creators of science's now ephemeral and disposable theories are currently being reduced to accord with their downgraded and devalued work, and with science's diminished ambitions."

ing their own campaign to save British philosophy, for the latter too has had its share of budget cuts⁶. In this article we argue that the British scientific and philosophical communities, and in particular the RS, have to their cost neglected one important factor in implementing their policies, and that therefore the present financial crisis is to a considerable extent self-inflicted.

On 17 and 22 February 1986 BBC television broadcast, in the highly regarded *Horizon* series, a film entitled "Science . . . Fiction?", and in the issue of 20 February 1986 *The Listener* published an article entitled "The Fallacy of Scientific Objectivity"⁷. As is evident from their titles, these were attacks against objectivity, truth and science. After rehashing the usual anti-science chestnuts, both the film and the article came to this conclusion: "The gradual recognition of these arguments may affect the practice, the *funding* and the institutions of science" (our italics). At least in Britain, the repercussions of these mistaken arguments are already happening. Scientists in other countries are duly forewarned.

We shall refer to these erroneous and harmful ideas as the epistemological anti-theses — the (un)philosophical positions which are contrary to the traditional and successful theses of natural philosophy. (Epistemology is the study of the nature, generation and validation of knowledge.)

Articles and programmes attacking the scientific theses and championing the anti-theses are published and broadcast regularly by the British media. But oddly, the RS, SBS and the other scientific bodies remain silent and do not deploy their powerful corporate muscle to answer such attacks against science (and sometimes against the RS itself). As a result, it appears, the editors of the popular media have come to the conclusion that the RS has no satisfactory answer.

This state of affairs is bad enough. But things are even worse: perversely, many individual scientists and philosophers seem bent on questioning and rejecting the true theses, and supporting the anti-theses. For example, most of the participants in the "Science . . . Fiction?" film were academic scientists. What is more, the RS apparently co-operated in the making of this programme: a scene in which the RS was explicitly assailed, was filmed in the RS's own London headquarters.

In order to demonstrate the threats that the anti-theses pose not only to science but also to society in general, it is useful to outline briefly the manner in which these ideas emerged gradually in the twentieth century.

In 1919 Sir Karl Popper by his own account⁸ had taken a strong dislike to the theories of Marx, Freud and Adler, whose supporters maintained that they were scientific. The difficulty was that Popper could not find any obvious way to refute them conclusively. Having noticed that Einstein's theories made (what seemed to him) falsifiable predictions, Popper resolved all his difficulties simply by declaring: "Irrefutability is not a virtue of a theory (as people often think) but a vice The criterion of the scientific status of a theory is its falsifiability"⁸. (Example: "The Earth is (approximately) a sphere" is not a scientific statement because it is not falsifiable; whereas "The Earth is a flat disk" is indeed scientific.) Popper also thought that observations are theory-laden. He phrased it thus: "Sense-data, untheoretical items of observation, simply do not exist. A scientific theory is an organ we develop outside our skin, while an organ is a theory we develop inside our skin"⁹.

But if observations are theory-laden, this means that observations are simply

theories, and then how can one theory falsify (never mind verify) another theory? Curiously, the full implications of this little complication were not fully grasped by Popper, but by Imre Lakatos: not only are scientific theories not verifiable, they are not falsifiable either — “If a theory is refuted, it is not necessarily false”¹⁰. (Example: both “the Earth is round” and “the Earth is flat” are neither verifiable nor falsifiable.) As David Stove¹¹ remarked, Lakatos has removed the logical stigma of falsity from refuted propositions, just as earlier public benefactors had removed the social stigma of illegitimacy from certain individuals.

So back to square one: if verifiability and falsifiability are not the criteria, then what makes a proposition scientific? It is hard to discern the answer to this question in Lakatos's writings. But if any answer is discerned at all, it is one that contradicts flagrantly the motto of the RS: “I am not bound to swear as any master dictates”¹⁴. This answer is more obvious in Thomas Kuhn's¹² writings: a proposition is scientific if it is sanctioned by the scientific establishment. (Example: if the scientific establishment decrees that “fairies exist”, then this would be scientific indeed.)

According to Kuhn, science is not the steady, cumulative acquisition of knowledge that was portrayed in old-fashioned textbooks. Rather, it is an endless succession of long peaceful periods which are violently interrupted by brief intellectual revolutions. During the peaceful period, which Kuhn calls “normal science”, scientists are guided by a set of theories, standards and methods, which Kuhn collectively designates as a “paradigm”. (Others call it a “world-view”.) During a revolution, the old paradigm is violently overthrown and replaced by a new one.

Vogues

So according to Kuhn, the business of science is not about truth and reality; rather, it is about transient vogues — ephemeral and disposable paradigms. In fact three pages from the end of his book *The Structure of Scientific Revolutions*¹², Kuhn himself drew attention to the fact that up to that point he had not once used the term “truth”. And when he used it, it was to dismiss it: “We may have to relinquish the notion that changes of paradigm carry scientists . . . closer and closer to the truth”. This passage was quoted approvingly by the author of a review in *Science*, whose final assessment of Kuhn's book was this:

Since Kuhn does not permit truth to be a criterion of scientific theories, he would presumably not claim his own theory to be true. But if causing a revolution is the hallmark of a superior paradigm, *The Structure of Scientific Revolutions* has been a resounding success¹³.

So now that the term “truth” has become a taboo, we know what must take

its place: mass popularity and prevailing fashion.

Kuhn's view, that a proposition is scientific if it is sanctioned by the scientific establishment, gives rise to the problematic question: what exactly makes an establishment “scientific”? This particular Gordian knot was cut by Paul Feyerabend: *any* proposition is scientific — “There is only one principle that can be defended under all circumstances and in all stages of human development. It is the principle: Anything goes”¹⁴. (Example: “There are one million fairies in Britain” is as scientific as “There are two hydrogen atoms in one water molecule”.)

In 1979 *Science* published a four-page complimentary feature¹⁵ about Feyerabend, the Salvador Dali of academic philosophy, and currently the worst enemy of science. In this article Feyerabend was quoted as stating that “normal science is a fairy tale” and that “equal time should be given to competing avenues of knowledge such as astrology, acupuncture, and witchcraft”. Oddly, religion was omitted. For according to Feyerabend (and the “Science . . . Fiction?” film too⁷), religion — and everything else — is an equally valid avenue of knowledge. In fact on one occasion Feyerabend characteristically put science on a par with “religion, prostitution and so on”¹⁴. This is a queer theme to pursue, but it is really a serious matter. For whereas in Britain the hot issue has been the cutbacks in the science budget, in the United States the most conspicuous impact of the antitheses has been the alarming growth of religious fundamentalism. All this and much else notwithstanding, the *Science* article's assessment of Feyerabend's monstrous ideas was: “Compared with the stiff and sober work that is often done in the philosophy of science, Feyerabend's views are a breath of fresh air”¹⁵.

Fortunately, the fallacies of the antitheses have been exposed by other academic philosophers, and the best debunking known to us was carried out by David Stove¹¹. Yet, one or other version of them remains popular with many philosophers, scientists, the media and the public.

The epistemological antitheses — scepticism, agnosticism, criticismism¹¹, cynicism, relativism, anarchism, nihilism — are not of course the invention of the twentieth century. Indeed the agnostic, nihilistic and anarchistic views of Gorgias, a fifth century BC Greek sophist, sound refreshingly modern. Aristotle¹⁶ (fourth century BC) reported that Gorgias taught that “nothing exists; and if anything exists, it is unknowable; and if it exists and it is knowable, yet it cannot be indicated to others”. Gorgias also argued that “one can persuade anyone of anything, if one speaks well enough”. The correct answer to this typical sophistry was already known to Aristotle: “One is more likely to

win one's argument, if what one says is true”. But it is sadly the case that those such as Gorgias and Feyerabend can fool a lot of people a lot of the time with sophistries like “anything goes”. But of course in the long run one thing goes — objective truth.

Falsity

For reasons known only to themselves, the advocates of the antitheses refuse to recognize that their ideas are flagrantly self-refuting — they negate and destroy themselves. And if a statement entails its own falsity, then it must be nonsense: the proposition “there is no truth” contradicts itself, and hence some truths must exist after all. Similarly, the stricture “nothing is certain, everything is doubtful” casts doubt, besides everything else, upon itself, and so some certain and undoubted things must exist. Likewise, if “anything goes”, then “nothing goes” must go too.

To most people this is sufficient refutation of the antitheses. But instead the sceptics and nihilists persist with doubting anything and everything with a cocksureness that beats any avowed dogmatist — they are dogmatical sceptics, and they would question anything except their own doubts. In the same vein, the epistemological relativists, who uphold the complete equivalence of all “paradigms”, may be aptly termed absolute relativists.

In any event, the spectacle of the supporters of the antitheses appealing to the very standards which they are bent on attacking is reassuring to those who uphold these standards. For example, those who claim to have refuted inductive reasoning have done so by employing inductive reasoning itself, as follows: from the true fact that many examples of inductive inference have indeed been shown to be invalid, these people inferred (inductively!) that all examples of inductive inference will be shown to be invalid.

On the subject of how to teach these weird modern ideas, we turn to the currently popular book *What Is This Thing Called Science?*¹⁷ by Alan Chalmers. It was first published in 1976 as a textbook for introductory university courses in the philosophy of science. Its popularity has been so great that it has been translated into many languages, and a new edition was published in 1982. The heading of the last section of the last chapter of the 1982 edition is “Why Bother?”, and the answer is revealing: “Why bother to carry out investigations of the kind to be found in the foregoing pages? The importance of the question becomes highlighted once it is admitted that philosophy or methodology of science is of no help to scientists”. After 168 pages packed with subtle concepts, sophisticated ideas and complex arguments, and after a gruelling academic session of (apparently futile) study, this must surely come as a rude

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Betrayers of the truth? Left to right: Karl Popper, Imre Lakatos, Thomas Kuhn and Paul Feyerabend.

shock to the serious student of science. But the student cannot complain. For Chalmers himself with commendable frankness notified the reader of the book from the outset, in the Introduction, what to expect from studying it: "We start off confused and end up confused on a higher level".

But come to think of it, once it is accepted, as done in this book, that the business of science is not about the discovery of truth and reality, then its conclusion that the methodology of science is of no help to scientists sounds eminently plausible: epistemological anarchism is a method, and nihilism is an (un)philosophical stance — and they do not require any special learning. Anyhow, Chalmers's answer to the hot question "Why Bother?", is that "as there is no timeless and universal conception of science or scientific method", the most important function of his book

is to combat what might be called the *ideology of science* as it functions in our society. This ideology involves the use of the dubious concept of science and the equally dubious concept of truth that is often associated with it, usually in the defence of conservative positions [italics in original].

While agreeing that genuine science cannot rationally defend any subjective ideological opinion, conservative or not, to dismiss science itself as just another ideology, and to allege that only passionate ideologues (but not dispassionate scientists) may benefit from the philosophy and methodology of science, is to talk nonsense. Yet sadly this is what many students of science, technology and medicine are taught in universities, and as a result these preposterous and dangerous ideas are becoming widespread within the scientific community. It is an objective of this article to refute these ideas, and argue that the correct epistemology is indispensable in any serious and responsible scientific work. For what is really at stake is nothing less than the future progress of our civilization.

One may wonder how many universities

in the world give their science students compulsory formal courses of lectures on the rigours of the scientific method. As for those universities which provide their students with an optional course on the current trends in the philosophy of science, are their governing bodies aware of the fact that many teachers of these courses are bent on sabotaging the scientific method?

Methodology

The hapless student is inevitably left to his or her own devices to pick up casually and randomly, from here and there, unorganized bits of the scientific method, as well as bits of *unscientific* methods. And when the student becomes the professional researcher, lacking proper tuition and training, he or she will grope haphazardly in the dark, follow expensive blind alleys and have recourse to such (unreliable) things as random guess, arbitrary conjecture, subjective hunch, casual intuition, raw instinct, crude imagination, pure chance, lucky accident and unplanned trial — and invariable error. Can this be an adequate methodology by means of which to make new discoveries and beneficial applications? Of course not, yet this is the entire methodology that the exponents of the antitheses actually recommend to professional researchers.

In the light of this evidence, Medawar's statement¹ does not seem surprising at all. In that passage Medawar suggested that the scientist "has no opinion" as to "what he conceives the scientific method to be". This might not be true of every scientist, but it appears to be true of quite a few. What is more, there has never been any attempt to rectify this worrying situation because, oddly and sadly, it was never seen as the anomalous circumstance that it indubitably is, and as the grave problem that it inevitably became.

Even though the endorsement of the antitheses saves one from the painstaking effort of discovering new truths, such conduct has some unfortunate drawbacks: the antitheses pose at least three different kinds of potential threats or actual

dangers not only to science as such, but also to society in general. In summary, these risks are: (i) Intellectual bankruptcy entails financial bankruptcy. (ii) Epistemological anarchism entails social anarchism. (iii) Epistemological relativism, criticismism¹¹ and nihilism entail scientific chaos, confusion and stagnation.

Danger (i). The first harmful repercussion of the antitheses, namely the cutback in the science budget, is already happening. The logical steps leading from the antitheses to the inescapable conclusion that the funding of science should be cut are not usually spelled out in detail. We spell them out now.

In the golden age of science, it was believed that the verified theories of science were true and everlasting. But with the increasing acceptance of the antitheses, these exalted old ambitions of science are seen as bogus. According to epistemological relativism, science should no longer claim superiority for its method and the knowledge that it produces. Moreover, in the capitalist system of economy, a scientific theory is a commodity like any other, and it must obey the laws of this system. Scientific theories are now considered to be temporary and dispensable. Furthermore, by denying truth and reality, the antitheses reduce science to a pointless, if entertaining, game; a meaningless, if exacting, exercise; and a destinationless, if enjoyable, journey. The aim of the game is just to play; the object of the exercise is merely to keep one busy; and the purpose of the journey is but aimless wandering.

Having lost their monopoly in the production of knowledge, scientists have also lost their privileged status in society. Thus the rewards to the creators of science's now ephemeral and disposable theories are currently being reduced to accord with their downgraded and devalued work, and with science's diminished ambitions.

It is the duty of those who want to save science, both intellectually and financially, to refute the antitheses, and to reassure their paymasters and the public

that the genuine theories of science are of permanent value, and that scientific research does have a concrete and positive and useful aim — to discover the truth and establish reality.

It is only on true knowledge that the socially beneficial and economically profitable medical and technological applications can be firmly grounded — knowledge normally discovered by means of the valid and judicious application of the scientific method. Now the hard fact is that most governments, being the political institutions that they are, provide public funds for research on the condition that the findings of this research will, sooner rather than later, have economic returns. For example, the British government told Parliament in 1986¹⁸ that it (the government) is

more likely to be persuaded of the value of increasing public investment in science, if the scientific community, and the users of its products, can point to increasing economic and social benefits, and in particular to prospects for increased national wealth.

Actually the ultimate reason cited by Shirley Williams in 1971 for the "dramatic decline in expenditure on scientific research in Britain", was that the "expected outcome" of the previously lavish spending, namely "an increase in gross national product", in fact "never happened"¹⁹.

In the wake of the rejection of scientific truth by so many researchers, and their candid admission that they do not really know what the scientific method is, the unpalatable fact pointed out by Williams is hardly surprising. If anything is surprising, it is the curious fact that some of the manufacturers of deficient, perishable and basically useless products — the short-lived, disposable and basically untrue theories — should complain that the selling price of their inferior product is dwindling.

It is therefore clear that the usual plea that greater funding of science results in proportionally greater wealth is woefully inadequate. Apart from increased funding, another requirement is increased care in the correct and sagacious application of the scientific method. Thus before laying all the blame at the doors of government and industry, the scientific community should put its own house in order.

It is not enough for scientists to assert vaguely that basic research will (just possibly) lead to new discoveries and profitable applications. Those scientists who want to uphold the intellectual integrity of their profession, and who are interested in selling their skills in the marketplace, owe it to their paymasters to explain adequately how they propose to make new and fruitful discoveries, instead of trying to conceal the embarrassing fact that they do not really know what the scientific method is.

Danger (ii). A second danger could have

an impact both broader and more disastrous than the first. This lies in the fact that epistemological cynicism and anarchism entail social, political and every other kind of anarchism and disorder. For it is not only in formal science that the issue of objective truth arises — it arises in every department of life. Both truths — ordinary and scientific — are of the same character in that they are both established by, either natural or artificial, either simple or

By denying truth and reality science is reduced to a pointless, if entertaining, game; a meaningless, if exacting, exercise; and a destinationless, if enjoyable, journey.

sophisticated, observations. In fact the only meaningful definition of truth is by way of the objectivity of theory-free and context-transcendent observation. If one kind of truth is denied, the other will have to go too. Now if the notion of objective truth established by observational evidence is disregarded, one is left in a chaos of arbitrary and conflicting opinions, opinions which are equally well- or ill-founded. For from the false premise that all observation is theory-dependent, all routes lead inexorably to "anything goes".

Danger (iii). A third danger springs from the fact that the antitheses are by nature obscurantist and they inevitably stifle progress. In the jungle of epistemological anarchism and criticism²⁰, every single bit of painstakingly proven knowledge is frivolously questioned and cynically scoffed at. In the consequent intellectual chaos there results a thick fog of confusion which completely obscures every possible route to further discoveries, and in this way the quest to enlarge knowledge is effectively paralysed.

Moreover, if one believes that there exists no objective truth, or even if one has doubts as to its existence, one then has no motivation, nor even inclination, to try to discover it. It is therefore highly unlikely that such a person would make any new discoveries (funding agencies take note). Furthermore, scepticism and nihilism were not threats only in the remote past; nor do they now constitute a risk of only academic interest. There seems to be evidence^{19,20} indicating that they may be impairing scientific progress at this moment.

Natural philosophy has had enemies throughout its 2,600 or so years of recorded history. But the present era is unique in that it is the first civilized society in which an effective anti-science movement flourishes contemporaneously with the unprecedentedly magnificent technological and medical applications of modern science. This is a curious paradox which cries out for clarification. The explanation of this sad situation is, of course, the large number of errors regard-

ing the nature, scope, method, powers and limitations of science.

It is not the objective of this article to correct all the errors which plague science and philosophy today, or to present the detailed solutions to all of the vexed problems of epistemology. For this is a colossal task, a task however that scientists are obliged to undertake themselves. The purpose of this article is merely to show that an acute problem exists, and to indicate broadly how it can be solved. The problem is that although the epistemological antitheses are demonstrably untenable, inherently obscurantist and positively dangerous, they have become alarmingly popular with the public, and even worse, with the communities of professional philosophers and scientists.

In barest outline, the solution of the problem is that science and philosophy will be saved — in both the intellectual and the financial sense — when the practitioners of these disciplines stop running down their own professions and start pleading the cause of science and philosophy correctly. This should be best done, first by thoroughly refuting the erroneous and harmful antitheses; secondly by putting forth adequate definitions of such fundamental concepts as objectivity, truth, rationality and the scientific method; and thirdly by putting the latter judiciously into fruitful practice. Only then will the expounding of the positive virtues of science and philosophy carry conviction. □

1. Medawar, P.B. *Induction and Intuition in Scientific Thought* 11 (Methuen, London, 1969).
2. Williams, S. *The Times* 15 (27 February 1971).
3. *The Public Understanding of Science* (The Royal Society, London, 1985).
4. *The Royal Society Corporate Plan: A Strategy for the Royal Society 1986-1996* (The Royal Society, London, 1986).
5. Porter, G. *Supplement to Royal Society News* 3, i-vi (1986).
6. Alexander, P. *et al. Analysis* 46, 161 (1986).
7. Lawson, H. *The Listener* 115, 12-13 (20 February 1986).
8. Popper, K.R. *Conjectures and Refutations: The Growth of Scientific Knowledge* 4th edn 33-37 (Routledge & Kegan Paul, London, 1972).
9. Popper, K.R. in *Problems in the Philosophy of Science* (eds Lakatos, I. & Musgrave, A.) 163-164 (North Holland, Amsterdam, 1968).
10. Lakatos, I. in *The Problem of Inductive Logic* (ed. Lakatos, I.) 397 (North Holland, Amsterdam, 1968).
11. Stove, D.C. *Popper and After: Four Modern Irrationalists* 16 (Pergamon, Oxford, 1982).
12. Kuhn, T.S. *The Structure of Scientific Revolutions* 2nd edn (University of Chicago Press, Chicago, 1970).
13. Wade, N. *Science* 197, 143-145 (1977).
14. Feyerabend, P. *Against Method: Outline of an Anarchistic Theory of Knowledge* 28 (New Left Books, London, 1975).
15. Broad, W.J. *Science* 206, 534-537 (1979).
16. Ross, W.D. (ed.) *The Works of Aristotle Volume VI Opuscula* 979^a 11-980^b 21 (Clarendon, Oxford, 1913).
17. Chalmers, A.F. *What Is This Thing Called Science? An Assessment of the Nature and the Status of Science and its Methods* 1st edn (University of Queensland Press, St Lucia, 1976); 2nd edn (Open University Press, Milton Keynes, 1982).
18. *The Future of the Science Budget: The Government's Response to the Report from the Education, Science & Arts Committee Session 1984-1985* (Command 9849, HMSO, London, 1986).
19. Sharp Cook, C. *Am. J. Phys.* 48, 175-176 (1980).
20. Psimopoulos, M. & Theocharis, T. *Am. J. Phys.* 54, 969 (1986).

T. Theocharis and M. Psimopoulos are in the Department of Physics, Imperial College of Science and Technology, London SW7 2AZ, UK.

HLA-DQ β gene contributes to susceptibility and resistance to insulin-dependent diabetes mellitus

John A. Todd, John I. Bell* & Hugh O. McDevitt

Departments of Medical Microbiology and Medicine, D-345 Fairchild Building, Stanford University School of Medicine, Stanford, California 94305, USA

Over half of the inherited predisposition to insulin-dependent diabetes mellitus maps to the region of chromosome 6 that contains the highly polymorphic HLA class II genes which determine immune responsiveness. Analysis of DNA sequences from diabetics indicates that alleles of HLA-DQ β determine both disease susceptibility and resistance, and that the structure of the DQ molecule, in particular residue 57 of the β -chain, specifies the autoimmune response against the insulin-producing islet cells.

EFFORTS to understand the mechanisms and genetics of complex polygenic human diseases have focused on the identification of DNA and protein markers that segregate with the disease in populations and families. In patients with some autoimmune disease¹, the most significant observation has been an increase in the frequency of certain serologically defined lymphoid cell-surface proteins encoded in the *HLA-D* region of the major histocompatibility complex (MHC; see Fig. 1) and referred to as class II antigens².

The HLA class II antigens are heterodimeric (α and β chains) glycoproteins that are normally expressed on the surface of B cells and antigen-presenting cells of the immune system². T helper cells can only recognize antigens (or their peptide fragments) that are associated with class II molecules. The interaction of the T-cell receptor, antigenic peptide and the class II molecule leads to T-cell activation and an immune response to the antigen³. It has been shown in the mouse⁴ and human⁵ that the primary sequence of the amino-terminal polymorphic domain is essential for T-cell recognition. Hence the immune responsiveness of an individual is partly determined by the polymorphic amino acids present in the class II molecules.

Insulin-dependent diabetes mellitus (IDDM, Type I) is the result of the selective destruction of the insulin-producing islet cells of the pancreas. It has been suggested that disease development involves an autoimmune response to an islet antigen(s) and is likely to be T-cell mediated⁶. The disease is clearly polygenic in inheritance⁷ and it has been estimated that the *HLA-D* region contributes over 50% of the heritability⁸. About

95% of IDDM patients possess either *HLA-DR3* or *-DR4*, compared to 45–54% of the normal population^{1,7}. At least one IDDM susceptibility gene has been shown in population^{7–10,12} and family^{11,13,14} studies to map to the MHC and is most tightly linked to the *HLA-D* region. MHC class II antigen serological^{15,16} and restriction fragment length polymorphism (RFLP; refs 15–23) studies have indicated that the *HLA-DQ* genes, which are in linkage disequilibrium with *HLA-DR*, are more strongly associated with IDDM than the *DR* genes. Despite this, no single determinant or RFLP has been found that is common to the several HLA haplotypes (*DR4*, 3 and 1) positively associated with this disease.

We have addressed this problem by sequencing the four major expressed polymorphic class II gene products isolated from three IDDM patients and several controls. To facilitate the sequencing of a relatively large number of HLA alleles, we have used a rapid method for producing sufficient quantities of the target gene sequences from RNA by complementary DNA synthesis and *in vitro* DNA amplification^{24,25}. Our main conclusions are: (1) there are no unique class II sequences found exclusively in IDDM patients, (2) the *DQ β* -chain amino-acid sequence is directly correlated with predisposition to IDDM and (3) this *DQ*-determined susceptibility is largely dependent on the identity of amino-acid residue 57 of the β -chain of this heterodimer. The murine homologue of *DQ β* , *A β* , isolated from the non-obese diabetic (NOD) mouse strain²⁶, when compared to other mouse *A β* alleles, has a unique Ser 57 residue that supports this correlation.

Sequencing class II alleles

We have cloned and sequenced the gene segments encoding the first domains of *DR β* , *DQ α* and *DQ β* chains that contain most

* Present address: University of Oxford, Nuffield Department of Clinical Medicine, John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK.

Fig. 1 Map of the class II loci of the human HLA-D region on the short arm of chromosome 6 (ref. 27). The *DR β* ₁ and *DR α* gene products associate to form a cell-surface heterodimeric glycoprotein. This molecule reacts with the alloantisera that define the major serological allotypes *HLA-DR1* to *DRw14*. The relatively non-polymorphic *DR β* _{III} and *DR α* gene products encode the *DRw52* and *DRw53* serological determinants, which are each associated with several *DR* allotypes. The genomic organizations of the *DR β* and *DR α* loci associated with the *DR3*, *DR4* and *DRw6* haplotypes are known^{43,44}. For the *DR2*, *DR5*, *DR7* and *DRw9* haplotypes we assume that the *DR β* ₁ gene encodes the more polymorphic chain. The *DR1* and *DRw8* haplotypes appear to have only one expressed *DR β* gene that corresponds to the *DR β* ₁ locus⁴⁵. On the *DR3* and *DR4* haplotypes^{43,44} the *DR β* _{II} gene is not expressed. The *DQ α* and *DQ β* genes encode the *DQ* serological specificities (*DQw1*, *DQw2* and *DQw3*). A fourth determinant called *DQ Blank*, has recently been defined by an alloantiserum, TK2 (ref. 29). The *DX α* and *DX β* and the second *DP α* and *DP β* genes are not expressed, and the *DO β* and *DZ α* genes are expressed only in very small amounts^{2,46–48}.

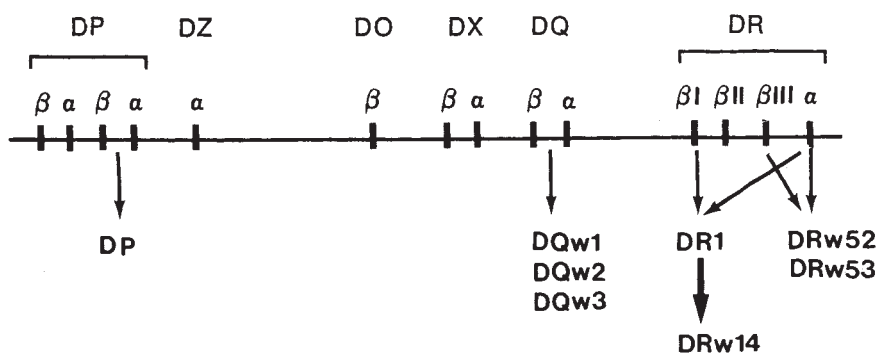


Table 1 Distribution of HLA haplotypes in healthy Caucasian populations and association of these with IDDM

DR	Dw	DQ	Haplotype frequencies (%)	IDDM association
1	1	w1.1	20	Positive
2	2	w1.2	26	Negative
2	12	w1.12	1.5	Neutral/negative
2	AZH	w1.AZH	1.5	Positive
3	3	w2	22	Positive
4	4	w3.1	5	Neutral/negative
4	4	w3.2	9	Positive
4	10	w3.2	3	Positive
4	14	w3.2	7	Positive
4	15	Blank	0.5	Neutral
5	5	w3.1	15	Neutral/negative
w6/w13	18	w1.18	10	Neutral/negative
w6/w13	19	w1.19	3	Positive
w6/w14	9	w1.9	3	Neutral
7	7	w2	27	Neutral
7		w3.2		
7		w3.3	1	
w8		Blank	6	Neutral
w8		w1		
		w3.1	2	
		w3.2		
w9		w3.2	1	Neutral
		w3.3		

Data are taken from several sources^{1,7-16,20,23,28-33}. For example, the DR4 antigen is found in 32% of the normal population. 95% of these DR4 chromosomes encode the DQw3 antigen which has two alleles, *DQw3.1* and *DQw3.2* of which 65% are *DQw3.2*. Cellular subtypes Dw10 and Dw14 are associated with DR4 chromosomes encoding the *DQw3.2* antigen but not with those encoding the *DQw3.1* antigen³¹. The DQ Blank antigen does not react with any of the standard antisera but has recently been characterised using an alloantiserum TK2 (ref. 29). The relative proportions of the *DQw3.2* allele and the novel *DQw3.3* allele (Fig. 3b) are currently unknown.

of the polymorphism. We synthesized cDNA from total RNA and amplified it *in vitro*^{24,25} using Klenow fragment and locus-specific oligonucleotide primers. Up to 750 base pairs (bp) can be specifically amplified, allowing rapid and efficient cloning and sequencing. We have found that amplification from cDNA has several advantages over genomic DNA: (1) complete sequences of the first domain are present, (2) pseudogenes that are not expressed are not amplified, and (3) there is more product, facilitating cloning. *In vitro* amplification using the commercially available, cloned and purified Klenow fragment is accurate and artifacts rare.

The genomic organization²⁷ of the currently identified class II loci is shown in Fig. 1 (the legend explains nomenclature and summarizes class II expression and serology). The disease associations of interest in this paper are with the class II alleles within the DR-DQ linkage group. In this group there are four polymorphic genes, *DR β 1*, *DR β III*, *DQ β* and *DQ α* . *DR α* is non-polymorphic² and the *DP* group is not analysed here because no significant association has been found between *DP* alleles and IDDM¹². The gene sequences from all four polymorphic loci on both chromosomes were obtained from three IDDM patients (two *DR3/4* heterozygotes and a *DR3* homozygote), together with the *DQ β* sequences from two other diabetics. The relevant polymorphic amino-acid sequences are presented for the first domains from the *DR β 1*, *DR β III* (Fig. 2) and *DQ α* (Fig. 3a) and *DQ β* (Fig. 3b) alleles isolated from these patients and nine healthy controls, which included several homozygous lymphoblastoid cell lines (HTC). For IDDM and the genes

analysed, all sequences found in patients were also found in healthy controls. Thus the IDDM autoimmune process is not due to mutant HLA class II alleles but is HLA-restricted by certain class II alleles of wild-type sequence which initiate normal T-cell responses. We have similar results for class II genes isolated from patients with multiple sclerosis and myasthenia gravis (unpublished data).

Position 57 of DQ β -chain

Alloantisera to DQ molecules define three major DQ antigens, DQw1, DQw2 and DQw3 (Fig. 1) (ref. 28). From the DQ α - and DQ β -sequences shown in Fig. 3 we can conclude that the DQw1 serological specificity is primarily determined by the α -chain, in agreement with previous results²⁵, and that the DQw2 and DQw3 specificities are determined by their respective β -chains. Each serologically defined DR antigen (DR1 to DRw14) is usually associated with one serologically defined DQw antigen (Table 1) (ref. 28). The use of antibodies^{15,16,29,30} and RFLP studies¹⁵⁻²³ have subdivided the DQw1 and DQw3 specificities into several alleles, but the DQw2-associated β -chain is invariant. Further complexity arises from the characterization of cellular subtypes, determined by stimulation in mixed lymphocyte reactions, and referred to as Dw (Table 1; ref. 31). Dw subtypes may be specified by either the DR β -chains or the DQ α - and β -chains, or both.

Table 1 shows the established distributions and major linkages of the various serologically-defined DR and DQ antigens and the Dw types in healthy Caucasian populations^{15,16,28-33} relevant to the present study. The associations of these antigens and subtypes with IDDM is also indicated. Many population studies have established that several class II haplotypes confer either increased or decreased risk for this disease^{1,7}.

The DR4 haplotype confers the highest relative risk for IDDM (refs 1, 7-14), even when different racial groups are compared^{9,14}. Two alleles of the *DQ β* locus have been identified which are associated with DR4, DQw3 positive haplotypes. One allele, designated *DQw3.2* (ref. 15), is found in ~90-95% of DR4 IDDM patients, compared to 60-75% of normal controls^{20,23,30,33}, and is negative for the serological determinant TA10 (refs 15, 16, 30, 33). The other allele, *DQw3.1*, is TA10 positive and its association with the IDDM is neutral or negative^{15,16,20,21}. Figure 3b shows that the *DQw3.1* allele DQ β -chain first domain sequence³⁴ has four amino-acid differences from the *DQw3.2* allele β -chain at positions 13, 26, 45 and 57. The DQ α -chains associated with both *DQw3* alleles have identical first domain amino-acid sequences (Fig. 3a).

We conclude that one or more of these four β -chain amino-acid substitutions are (1) probably responsible for TA10 reactivity and may therefore influence the three-dimensional structure of the heterodimer and (2) involved in the positive association with IDDM of the *DQw3.2* allele relative to the *DQw3.1* allele. Analysis of these and other DQ β -chain polymorphic residues in Table 2 shows that only the amino acid at position 57 of the DQ β -chain is strongly correlated with susceptibility and resistance to IDDM. All of the *DQ β* alleles that are positively associated with IDDM (*DR4-DQw3.2*, *DR3-DRw2*, *DR1-DQw1.1*, *DR2-DQw1.AZH*) have either Ala, Val or Ser at position 57. Asp is present at position 57 in all the *DQ β* alleles which are neutral or negatively associated with disease (*DR4-DQw3.1*; *DR2-DQw1.2*; *DR2-DQw1.12*).

The DR2 haplotype (Table 1) provides a striking example of this correlation. Although the subtype AZH accounts for only about 6% of normal DR2 haplotypes³¹, 77% of IDDM DR2 chromosomes are DR2-AZH (refs 20, 22, 35). The positive associations of the DR1 and DR2-AZH haplotypes are consistent with the structures of their DQ β chains, which are identical except for the Val-Ser difference at position 57. Data from an oligonucleotide dot-blot analysis of Caucasian IDDM patients and the NOD mouse support this correlation (see below).

	10	20	30	40	50	60	70	80	90
DR1β1	G D T R P R F L V Q	L K F E C H F F M G	T E R V R L L E R C	I Y N Q E E S V R F	D S D V G E Y R A V	T E L G R P D A E Y	W N S Q K D L L E Q	R R R A V D T Y C R	H N Y G V G E S F T V Q R R
DR2 Dw2β1	---	D - Y - - - - -	F - H - D - - - -	---	D L - - - - -	---	F - D - - - -	---	---
DR2 Dw2β111	---	P - R - - - - -	F - D - Y - - - -	---	F - - - - -	---	I - A - - - -	---	---
DR2 Dw12β1	---	D - Y - - - - -	F - H - G - - - -	---	M - - - - -	---	F - D - - - -	---	---
DR2 Dw12β111	---	P - R - - - - -	F - D - Y - - - -	---	F - - - - -	---	I - A - - - -	---	---
DR2 AZHβ1	---	C - G - - - - -	D - Y - - - - -	F - H - G - - - -	---	N - - - - -	---	I - A - - - -	---
DR2 AZHβ111	---	P - R - - - - -	F - D - Y - - - -	---	F - - - - -	---	F - D - - - -	---	---
DR3β1	---	E Y S T S - - - -	---	Y D Y F H - - - -	---	M - - - -	---	K - G R - H - - - -	---
DR3 DRw52a	---	E L R - S - - - -	---	Y D Y F H - - - -	---	F L - - - -	---	K - G R - H - - - -	---
DR3 DRw52b	---	E L - S - - - -	---	F - H - - - -	---	Y A - - - -	---	R - G G - H - - - -	---
DR4 Dw4β1	---	E - - - - -	---	F - D - Y - - - -	---	F H - - - -	---	R - - - - -	---
DR4 Dw10β1	---	E - - - - -	---	F - D - Y - - - -	---	F H - - - -	---	I - D - - - -	---
DR4 Dw14β1	---	E - - - - -	---	F - D - Y - - - -	---	F H - - - -	---	I - D - - - -	---
DR4 Dw53	---	O - - - - -	---	E A C - - - -	---	L - - - -	---	M - - - -	---
DR5 Dw5β1	---	E Y S T S - - - -	---	F - D - Y - - - -	---	F H - - - -	---	I - D - - - -	---
DRw6β1	---	E Y S T S - - - -	---	F - D - Y - - - -	---	F H - - - -	---	I - D - - - -	---
DRw6β1	---	E Y S T S - - - -	---	F - D - Y - - - -	---	F H - - - -	---	I - D - - - -	---
DR7β1	---	O - - - - -	---	G Y K - - - -	---	Q F - L - - - -	---	F - - - -	---
DRw8β1	---	E Y S T G - - - -	---	F - D - Y - - - -	---	F H - - - -	---	I - D - - - -	---
DRw9β1	---	O - - - - -	---	K - - - -	---	D - - - -	---	Y - H - G - - - -	---

Fig. 2 DR β -chain first domain amino-acid sequences from the following cell lines, HTCs and lymphoblastoid cell lines and peripheral blood lymphocytes (PBL) from healthy and diseased subjects. DR1 β I: LG2 (HTC) (ref. 45), B-cell line 45.1 (ref. 41). DR2-Dw2 β 1: PGF (HTC) (ref. 45). DR2-Dw2 β 111: PGF (ref. 45). DR2-Dw12 β 1: BGE (HTC) (ref. 49), DHO (HTC) (ref. 50). DR2-Dw12 β 111: BGE (HTC) (ref. 49). DR2-AZH β 1 (ref. 49) (or DR2-MN2) (ref. 50): AZH (HTC) (ref. 49), individual types as DR2-MN2/DR4-Dw4 (ref. 50). DR2-AZH β 111: AZH (ref. 49), DR2-MN2/DR4-Dw4 (ref. 50). DR3 β 1: AVL (HTC) (ref. 51), WT49 (HTC) (ref. 45), DM24 (DR3/4, IDDM), DM28 (DR3/3, IDDM), DM29 (DR3/4, IDDM). DRw52a (DR3 β 111): AVL (ref. 51), DM28, DM29. DRw52b: DR4/w6 cell line (ref. 41), DM24. DR4-Dw4 β 1: Priess (HTC) (ref. 43), GM2103 (HTC) (ref. 39), WT51 (HTC) (ref. 45). DR4-Dw10 β 1: FS (HTC) (ref. 39), DM24. DR4-Dw14 β 1: BIN-40 (HTC) (ref. 39), DM29. DR4 DRw53: Priess (ref. 43), DM24, DM29 (see also identical DRw53 sequences from DR7 and DRw9 HTCs (ref. 45)). DR5 Dw5 β 1: Swei (HTC) (ref. 52), DRw6a β 1: HHK (HTC) (ref. 51), APD (HTC) (ref. 45). DR6b β 1: DR4/w6 cell line (ref. 51). DR7 β 1: LBF (HTC) (ref. 45). DRw8 β 1: Madura (HTC) (ref. 45). DRw9 β 1: DKB (HTC) (ref. 45). See also two DR β gene sequences from an IDDM patient⁵³ typed as DR4-Dw4/DR2-AZH. The numbers above the sequence refer to the amino-acid positions. A dash indicates identity with the DR1 sequence and blanks indicate that no sequence information is available. Nucleotide sequences will be supplied on written request. Cell lines and PBL listed without reference were sequenced as part of this study.

Methods. RNA was purified from established lymphoblastoid cell lines or separated PBL from 50–100 ml of blood and cDNA synthesized⁵⁴ from 15–20 μ g of total cellular RNA. cDNA was subjected to two sets of 15 cycles of DNA polymerase I Klenow fragment-catalysed *in vitro* DNA amplification²⁵. The annealing and extension reactions were at 37 °C and the polymerase buffer²⁵ contained no NaCl and 5 mM MgCl₂. After 15 cycles, the amplified products were purified by organic extractions and two ammonium acetate-ethanol precipitations. The final pellet was resuspended in 50 μ l of H₂O for the next 15 cycles. This procedure produced more efficient amplification than a continuous set of 28–30 cycles. The DR β oligonucleotide primers are: DR β LP1 (5'-ATGGTGTGTCTGAAGCTCCC-3') anneals to the region encoding the amino acids -29 to -23 at the beginning of the leader peptide and DR β AMP1 (5'-CGCTGCACTGTGAAGCTCTC-3') is complementary to the RNA and anneals to the region encoding amino acids 87–93 (ref. 41). The primers were designed to amplify all the published DR β alleles but not DQ β . The specific amplified product (about 365 bp) was visualized on an ethidium bromide-stained low-melting agarose gel and purified by electroelution on to (Schleicher and Schüll) NA45 membrane, followed by three organic extractions and two ethanol precipitations. About 10% of this gel-purified product routinely gave 100–200 M13 plaques after ligation (T4 ligase, Biolabs) and transformation into *Escherichia coli* JM101. The M13 vector was identical to M13K8.2 (ref. 55) except that the *Sma*I site has been replaced by an *Eco*RV site (P. Carter, personal communication). 70%–90% of these plaques contained class II sequences and hence were dideoxy-sequenced directly. More than one clone was isolated for each allele and sequenced on both strands.

Table 2 The association of IDDM with DQ β polymorphic residues

DR antigen	DQ β allele	IDDM association	3	9	13	14	26	28	30	37	38	45	46	47	52	53	55	56	57	66	67	70	71	74	75	77	84	85	86	87	89	90				
DR4	w3.2	Positive	S	Y	G	M	L	T	Y	Y	A	G	V	Y	P	L	P	P	A	E	V	R	T	E	L	T	Q	L	E	L	T	T				
DR3	w2	Positive	-	-	-	-	-	S	S	I	V	-	E	F	L	-	L	-	-	D	L	-	K	A	V	R	-	-	-	-	-	-				
DR1	w1.1	Positive	-	-	-	L	G	-	H	-	V	-	-	-	-	Q	R	-	V	-	-	G	A	S	V	R	E	V	A	Y	G	I				
DR2	w1.AZH	Positive	-	-	-	L	G	-	H	-	V	-	-	-	-	Q	R	-	S	-	-	G	A	S	V	R	E	V	A	Y	G	I				
DR4	w3.1	Neutral/negative	-	-	A	-	Y	-	-	-	-	E	-	-	-	-	-	-	D	-	-	-	-	-	-	-	-	-	-	-	-	-				
DR5	w3.1	Neutral/negative	-	-	A	-	Y	-	-	-	-	E	-	-	-	-	-	-	D	-	-	-	-	-	-	-	-	-	-	-	-	-				
DR2	w1.12	Neutral/negative	P	L	A	-	Y	-	-	D	V	-	-	-	-	Q	R	-	D	D	I	-	-	-	-	-	E	V	A	F	G	I				
DR2	w1.2	Negative	-	F	-	-	-	-	-	-	-	-	-	-	-	Q	R	-	D	-	-	G	-	-	-	-	E	V	A	F	G	I				
Mouse																																				
NOD	A β NOD	Positive	-	H	-	E	-	-	-	-	L	-	E	-	E	-	R	H	S	Y	*	-	-	-	-	-	E	E	T	E	P	-				
B10	A β (b)	Negative	-	Y	-	E	Y	-	-	-	V	-	E	H	E	-	R	-	D	E	I	-	-	-	-	-	E	G	P	E	H	-				

A comparison of all the polymorphic residues of the DQ β alleles associated with DR haplotypes that have been studied in IDDM and control populations. The assignments of IDDM association and order are based on the ranking system previously described⁷, on data discussed and cited in the text, and Figs 2 and 3. The strength of the association (negative, neutral or positive) is to some extent dependent on amino acid 57 but also on other amino acids of the β -chain and the α -chain of the DQ molecule. The asterisk in the NOD mouse A β sequence denotes a deleted amino acid. A dash indicates identity with the DQw3.2 sequence.

<i>a</i>	10	20	30	40	50	60	70	80
DR1 DQw1.1	EDIVADHVAS	CGVNLVQFYG	PSGQYTHEFD	GDEEFYVDLE	RKETAWRWPE	FSKFGGFDPPQ	GALRNHMAVAK	HNLHIMIKRY
DR2 DQw1.2	---	---	---	Q	---	---	---	NSTAATN
DR2 DQw1.2ZH	---	---	---	Q	---	---	---	---
DR3 DQw2	---	Y	S	---	Q	---	---	---
DR4 DQw3.1	---	Y	S	---	---	V	CL	V
DR4 DQw3.2	---	Y	S	---	---	V	CL	V
DR5 DQw3.1	---	Y	S	---	---	V	CL	V
DR6 DQw1.18	---	---	---	---	---	---	---	---
DR6 DQw1.19	---	---	---	---	---	---	---	---
DR7 DQw2 and w3	---	Y	---	S	---	---	---	---
DR8 DQw1	---	---	---	---	---	---	---	---
DR8 DQw3	---	Y	S	---	---	---	---	---
DR9 DQw3	---	Y	S	---	---	---	---	---

<i>b</i>	10	20	30	40	50	60	70	80	90
DR1 DQw1.1	RDSPEDFVYQ	FKGLCYFTNG	TERVRGVTNRH	IYNREETVRF	DSVDGVYRAV	TPOGRPVAEY	WNSQKEVLEG	ARASVDRVCR	HNYEVAYRGI
DR2 DQw1.12	---	---	---	---	---	---	---	---	---
DR2 DQw1.2	---	---	---	---	---	---	---	---	---
DR2 DQw1.2ZH	---	---	---	---	---	---	---	---	---
DR3 DQw2	---	---	---	---	---	---	---	---	---
DR4 DQw3.1	---	---	---	---	---	---	---	---	---
DR4 DQw3.2	---	---	---	---	---	---	---	---	---
DR4 DQ BLANK	---	---	---	---	---	---	---	---	---
DR5 DQw3.1	---	---	---	---	---	---	---	---	---
DR6 DQw1.9	---	---	---	---	---	---	---	---	---
DR6 DQw1.18	---	---	---	---	---	---	---	---	---
DR6 DQw1.19	---	---	---	---	---	---	---	---	---
DR7 DQw2	---	---	---	---	---	---	---	---	---
DR8 DQ BLANK	---	---	---	---	---	---	---	---	---
DR9 DQw3.3	---	---	---	---	---	---	---	---	---

Fig. 3 *a*, DQ α -chain first domain amino-acid sequences from the following cell lines, HTCs and lymphoblastoid cell lines and peripheral blood lymphocytes from healthy and diseased subjects. DR1 DQw1.1: LG2 (ref. 25) BML (healthy control typed as DR1/4, DRw53, DQw1, DQw3 and TA10 positive); DR2 DQw1.2: PGF (ref. 25); DR2 DQw1.2ZH: DRA AZH; DR3 DQw2: Raji (ref. 56), DM24, DM28, DM29; DR4 DQw3.1: BML; DR4 DQw3.2: JY cell line (DR4/w6) (ref. 46), DM24, DM29; DR5 DQw3.1: MG3 (myasthenia gravis patient typed as DR5/w9); DRw6 DQw1.18 WVB (HTC); DRw6 DQw1.19: WT46 (HTC); DR7 DQw2: LG-10 (HTC) (ref. 36); DR7 DQw3: Beiler (HTC) (R. Karr, personal communication); DRw8 DQ BLANK: Madura (HTC); DRw8 DQw1: TAB (HTC) (ref. 25); DRw9 DQw3.3: DKB (HTC), ISK cell line (ref. 36). (ISK, homozygous for DRw9, has an identical sequence to the DKB sequence except that it contains a five-amino-acid deletion from position -1 to +4 and has one silent base pair substitution in the codon for amino-acid position 7.) The numbers refer to the amino-acid residues. A dash indicates identity with the DR1 sequence and a blank indicates that no sequence in formation is available. Nucleotide sequences will be supplied on request for both DQ α and DQ β . *b*, DQ β -chain first domain amino-acid sequences from the following sources: DR1 DQw1.1: B-cell line 45.1 (ref. 48), BML, MVL (HTC); DR2 DQw1.2: PGF; DR2 DQw1.12: BGE (ref. 57); DR2 DQw1.2ZH: DR2 AZH (ref. 57); DR3 DQw2: WT49 (HTC) (ref. 58), DM24, DM28, DM29; DR4 DQw3.2: DR4 homozygous individual⁴⁰, 3102 (Dw4) (ref. 39), FS (Dw10) (ref. 39), BIN 40 (Dw14) (ref. 39), DM24 (Dw10), DM29 (Dw14); DR4 DQw3.1: BML, DM23 (DR4, w6 IDDM patient); DR4 DQ BLANK: KT3 (Dw15; HTC) (ref. 39); DR5 DQw3.1: MG3; DRw6 DQw1.9: WT52 (HTC); DRw6 DQw1.18 WVB (HTC); DKB (DR4, w6 IDDM patient); DRw6 DQw1.19: Daudi (DRw6 cell line)⁶⁰; DM23 (DR4, w6 IDDM patient); DR7 DQw2: Burkhardt (HTC) (ref. 49); DRw8 DQ BLANK: Madura; DRw9 DQw3.3: DKB. The second number in the DQw designations (for example, 18 in DQw1.18) refers to the Dw cellular typing number associated with the DQ allele.

Methods. See also Fig. 2 legend. For DQ α the oligonucleotide primers for cDNA amplification are: LPDQ α , 5'-CTGACCAC-CGTGATGAGCCC-3' (amino-acid positions -10 to -4, according to numbering system in ref. 56) and AMPDQ α , 5'-GGTGAG-GTAACTGATCTTGAAG-3' (amino-acid positions 147-154) whose product is ~490 bp long. The primers for DQ β cDNA amplification are LPDQ β , 5'-GAGCAGCCAGTGGCTGAGGG-3' (amino-acid positions -8 to -1, according to numbering system in ref. 40) and AMPDQ β , 5'-GTTGTGGTGGTTGAGGGCCTC-3' (amino-acid positions 107-113) with ~360 bp-long product. The DRw8-DQ Blank DQ β sequence was generated from poly(A)⁺ RNA (kindly provided by D. Denney).

Oligonucleotide dot blot analysis

Table 3a shows the distribution of serologically-defined HLA-DR4, 3 and 2 antigens in 39 randomly selected, Caucasian IDDM patients, compared to the published antigen frequencies in healthy Caucasian controls (Table 1). In agreement with previous studies^{1,7}, 95% of the patients carry DR3 and/or DR4. The next most common antigen is DR1 (data not shown). A significant positive association of DR1 with IDDM has been reported⁷. DR2 is absent from our IDDM population. Significant negative associations of DR2 have been found consistently with IDDM, including studies involving different racial groups^{1,7,14}. Only two of our patients carry the DR5 antigen (data not shown). The DR5 haplotype is in linkage disequilibrium with the DQw3.1 allele²⁸ and has been negatively associated with IDDM (ref. 14).

Specific DQ β alleles can be identified in patients after *in vitro* amplification of genomic DNA by using allele-specific oligonucleotides to probe dot blots of the amplified materials. This rapid and sensitive approach has been described for the DQ α alleles²⁵. Using one or two oligonucleotide probes for each of the DQ β alleles DQw3.1, DQw3.2, DQw2 and DQw1.1 we have genotyped our panel of 39 random IDDM patients and a number of healthy Caucasian controls.

An example of the primary data is shown in Fig. 4 and the results for DQw3.1 and 3.2 are given in Table 3b. The probes and conditions used are described in the legend to Table 3. The DQw2 and 1.1 probes gave an exact match with DR3 and DR1 serotypes.

In agreement with previous studies^{18-21,23} 25 out of 27 (93%) DR4 IDDM chromosomes were DQw3.2 positive and DQw3.1 negative compared to 16 out of 21 (76%) DR4 healthy, Caucasian chromosomes ($P=0.12$, Fischer's exact test; P values are uncorrected). Strikingly, the two non-DR3, non-DR4 patients (DR7/- and 6/-) carried the DQw3.2 allele. Of our 39 diabetics, 80% carry at least one copy of the DQw3.2 β -chain allele. For the DRw6 associated haplotypes, the data (not shown) suggest that the DQw1.19 allele (Val 57; see Fig. 3b) is positively associated with IDDM. It is noted that the DR7 haplotype usually carries the susceptibility allele, DQw2, and yet is not increased in frequency in patients⁷. This might be due to the unique structure of the DR7 DQ α -chain³⁶.

NOD mouse class II sequence

Our observed correlation is supported by inspection of the A β first domain sequence isolated from the NOD mouse (Table 2 and ref. 26). This spontaneous disease is autoimmune in nature

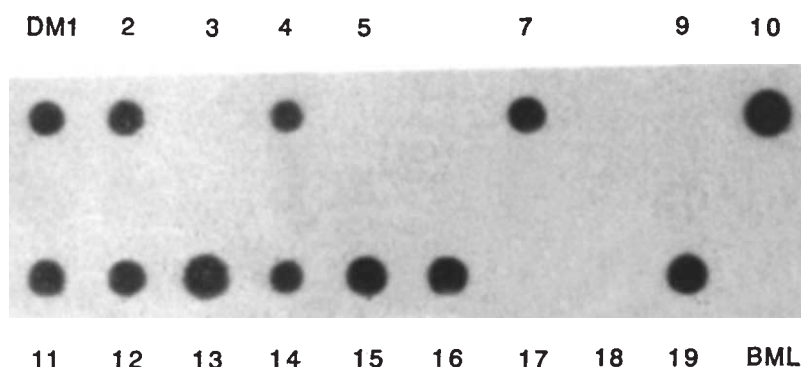


Fig. 4 Oligonucleotide dot blot analysis of 17 IDDM patients (DM1-19) and one control, BML. As described in the Table 3 legend, *DQ_β* locus-amplified genomic DNA was blotted on to nitrocellulose, probed with oligonucleotide DQw3.2, washed at 65 °C in 6 × SSC/0.5% SDS and autoradiographed for 24 h. The negative healthy control, BML, is DR1, 4 and TA10/DQw3.1 positive (see sequences in Fig. 3) and is negative for the DQw3.2 oligonucleotide. Patients DM3, 5, 9, 17 and 18 were also negative.

and is similar to human IDDM in both genetics and pathogenesis⁶. Crosses of NOD with a normal mouse strain (C57BL/10, which is also I-E negative) showed that the development of diabetes was dependent on homozygosity for the NOD MHC³⁷. The first domain sequence of the NOD mouse *A_β* chain, which is the homologue of the human *DQ_β*-chain, is unique among murine *A_β* alleles in that it has a Ser residue at position 57. All other published mouse *A_β* alleles have Asp at position 57³⁸. These data indicate that the NOD *A_β* gene shares the same susceptibility determinant as human *DQ_β* and most probably contributes to disease susceptibility.

HLA-DR allele associations

RFLP analysis of IDDM patients has shown that certain *DR_β* polymorphisms subdividing HLA-DR3 (refs 18–20, 23) occur at a higher frequency than in controls. We have analysed these associations in our IDDM panel by using a *DR_β* allele-specific oligonucleotide and show that, for this small sample, one of these associations is not significant. The *DRw52a* allele was detected in 18 of 23 *DR3* IDDM patients versus 8 of 13 healthy *DR3* controls ($P = 1.81$).

Several cellular subtypes of *DR4* have been described³¹ (Table 1) and three of them, Dw4, Dw10 and Dw14, have been attributed to a small number of amino-acid substitutions in the *DRβ1* chain^{5,39} (Fig. 2). Dw10-specific oligonucleotide analysis of the IDDM patients showed no significant difference between patients and controls (4 Dw10 out of 27 *DR4* IDDM patients versus 3 Dw10 out of 19 *DR4* controls; $P = 0.33$). We conclude that although the *DRβ1* and *DRβ11* loci could influence IDDM susceptibility, they do not do so to the same extent as the *DQ_β* locus.

Conclusions

Our data indicate that *DQ_β* alleles that do not have Asp at position 57 are significantly enriched in the Caucasian IDDM population. We propose that the amino acid at position 57 determines a critical function of the *DQ* molecule in IDDM. We do not know at present whether Asp-57 alone will confer decreased susceptibility. This, with other polymorphic and conserved residues, must contribute to the overall structure of the *DQ* molecule. We suggest that *DQ_β* allelic polymorphisms, particularly at position 57, determine the specificity and extent of the autoimmune response against islet cell antigens through T cell help and/or suppression. Full HLA-susceptibility is therefore dependent on the individual having two Asp-57-negative *DQ_β* alleles, especially if they are from the *DR4* and/or *DR3* haplotypes. If the individual carries one Asp-57-negative allele and one Asp-57-positive allele (*DQw1.2* or *DQw3.1*) then that person has a much lower risk of developing IDDM. Of our 39 IDDM patients, 10% are heterozygous for *DQ_β* Asp 57. *DQ_β* Asp 57-negative homozygosity is thus necessary (but not sufficient) for disease development in 90% of Caucasian Type I diabetics. The possession of two Asp 57-positive *DQ_β* alleles provides almost complete resistance to IDDM.

The observed requirement for homozygosity of Asp 57-negative *DQ_β* alleles could explain the apparent recessive inheritance of MHC-linked susceptibility to IDDM seen in man¹³ and the NOD mouse³⁷. Recessive inheritance is unusual in HLA-class-II-associated susceptibility to autoimmune disease¹, and sug-

Table 3 Serological and allele-specific oligonucleotide dot blot analysis of 39 IDDM patients

a		
HLA-DR antigen	Number of patients (n = 39)	Healthy controls (%) [*]
DR4	28 (72%)	32
DR3	25 (64%)	22
DR3, 4	17 (44%)	3
DR2	0	26
b		
<i>DQ_β</i> allele	Patient chromosomes (%)	Control chromosomes (%)
<i>DR4-DQw3.1</i>	7 (2/27)	24 (5/21)
<i>DR4-DQw3.2</i>	93 (25/27)	76 (16/21)

The 39 IDDM patients were all Caucasian, most with early disease onset (age less than 20 years) and continuously treated with insulin since diagnosis. Genomic DNA (1–2 μg) was subjected to *in vitro* amplification and probed with allele-specific oligonucleotides by dot blot analysis²⁵. The genomic DNA *DQ_β* primers amplify *DQ_β* and *DX_β* loci and anneal to sequences within exon 2 of these genes. They are GLPDQβ1, 5'-GATTCGTGTACCAAGTTAAGGGC-3' (amino-acid positions 6–13, ref. 40) and GAMPDQXβ2, 5'-CCACCTCGTAGT-TGTGTCTGCA-3' (amino-acid positions 79 to 86; product size ~240 bp). The genomic DNA *DR_β* primers are GLPDRβ1, 5'-TTCTTCAATGGGACGGAGCG-3' (amino-acid positions 17–23, ref. 41) and GAMPDRβ1 5'-GCCGCTGCACTGTGAAGCTCTC-3' (amino-acid positions 87–94; product size ~230 bp). After 28 cycles of amplification, 20% of the sample was applied to nitrocellulose with a BRL Hybri-Dot manifold. Dot blots were probed⁴² at 42 °C with allele-specific oligonucleotides; after hybridization, blots were washed in 6 × SSC/0.1% SDS at 30 °C for 10 min and autoradiographed for 1–4 h to assess the degree of hybridization. Blots were then washed in 6 × SSC/0.5% SDS at a temperature based on: $T_d = (\text{number of GC bps}) \times 4 + (\text{number of AT bps}) \times 2$. This removed probe that had one mismatch or more with the target sequence. The following oligonucleotide probes were used to identify specific alleles (the indicated temperatures are those required to wash off mismatched probe): DQw3.2, 5'-GGCCGCTGCGCCGAG-3' (amino-acid positions 54–59, Fig. 3b) at 65 °C; DQw3.1-57 5'-GGCCGCTGACGCCGAG-3' (amino-acid positions 54–59) at 67 °C and DQw3.1-26 5'-CGTGCGTTATGTGACCA-3' (amino-acid positions 23–29) at 52 °C; DQw1.1, 5'-GGCGGCTGTGCGGAG-3' (amino-acid positions 54–59) at 62 °C; DQw2, 5'-GCTGGGGCTGCGTCCG-3' (amino-acid positions 52–58) at 62 °C; DRw52a, 5'-CAGGAGGAGTTCCTGCGC-3' (amino-acid positions 34–39, Fig. 2) at 60 °C; Dw10, 5'-GAAGACGAGCGGGCCGCG-3' (amino-acid positions 69–74) at 64 °C (see text for data).

^{*} See Table 1.

gests a distinct immunological mechanism. One explanation for dominant non-responsiveness is cross-tolerance between a self-antigen and the antigen under study. An alternative mechanism is MHC class II-restricted immune suppression.

The *DR2-Dw2 DQ β* allele is unique in that a single dose of this gene provides strong protection against IDDM. This is not complete, as a few IDDM patients (less than 0.5% total patients) have been typed as *DR2-Dw2* (refs 20, 22, 35). The mechanism of DR2-mediated IDDM protection is unknown but is clearly dominant. It is possible that the unique structure of the *DR2-Dw2 DQ β* -chain provides cross-tolerance to the self antigen.

We thank H. Acha-Orbea, P. Carter, H.-J. Garchon, S.

Hughes, K. Karagianes, A. McMichael, K. Moody, S. Todd, P. Travers, J. Trowsdale and D. Wraith. We also thank D. Denney, M. Greenfield, C. Grumet, F. Kraemer, C. Pask, P. Patek, R. Payne, T. Sasazuki, R. Spielman and L. Steinman for reagents and material from patients, and H. Acha-Orbea, R. Karr, B. Lee, M. Segall, R. Spielman and M. Trucco for unpublished data. This work was funded by the European Molecular Biology Organisation and the Science and Engineering Research Council (J.A.T.), the Alberta Heritage Fund for Medical Research (J.I.B.) and by National Institutes of Health.

Note added in proof: Similar results have been obtained by H. Erlich and G. Horn (personal communication).

Received 15 June; accepted 8 September 1987.

- Svejgaard, A., Platz, P. & Ryder, L. P. *Immun. Rev.* **70**, 193-218 (1983).
- Kaufman, J. F., Auffray, C., Korman, A. J., Shackelford, D. A. & Strominger, J. L. *Cell* **36**, 1-13 (1984).
- Sette, A., Buus, S., Colon, S., Smith, J. A., Miles, C. & Grey, H. *Nature* **328**, 395-399 (1987).
- Germain, R. N. & Malissen, B. A. *Rev. Immun.* **4**, 281-315 (1986).
- Cairns, J. S. *et al. Nature* **317**, 166-168 (1985).
- Rossini, A. A., Mordes, J. P. & Like, A. A. *Rev. Immun.* **3**, 289-320 (1985).
- Thomson, G. *Am. J. hum. Genet.* **36**, 1309-1317 (1984).
- Rotter, J. I. & Landlaw, E. M. *Clin. Genet.* **26**, 529-542 (1984).
- Winter W. E. *et al. New Engl. J. Med.* **316**, 285-291 (1987).
- Rich S. S., Weitkamp, L. R. & Barbosa, J. *Am. J. hum. Genet.* **36**, 1015-1023 (1984).
- Winearls, B. C. *et al. Tissue Antigens* **24**, 234-246 (1984).
- De Jongh, B. M., Termijtelen, A., Bruining, G. J., De Vries, R. R. P. & Van Rood, J. J. *Tissue Antigens* **23**, 87-93 (1984).
- Rubinstein, P., Ginsberg-Fellner, F. & Falk, C. *Am. J. hum. Genet.* **33**, 865-882 (1981).
- Bertrams, J. & Baur, M. P. in *Histocompatibility Testing* (eds Albert, E. D., Baur, M. P. & Mayr, W. R.) 348-358 (Springer, Berlin, 1984).
- Kim, S. J. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 8139-8143 (1985).
- Schreuder, G. M. T. *et al. J. exp. Med.* **164**, 938-943 (1986).
- Owerbach, D. *et al. Nature* **303**, 815-817 (1983).
- Cohen-Haguenauer, O. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 3335-3339 (1985).
- Arnheim, N., Strange, C. & Erlich, H. *Proc. natn. Acad. Sci. U.S.A.* **82**, 6970-6974 (1985).
- Bohme, J. *et al. J. Immunol.* **137**, 941-947 (1986).
- Nepom, B. S. *et al. J. exp. Med.* **164**, 345-350 (1986).
- Segall, M. *et al. Hum. Immunol.* **17**, 61-68 (1986).
- Festenstein, H. *et al. Nature* **322**, 64-67 (1986).
- Scharf, S. J., Horn, G. T. & Erlich, H. *Science* **233**, 1076-1078 (1986).
- Saiki, R. K., Bugawan, T. L., Horn, G. T., Mullis, K. B. & Erlich, H. A. *Nature* **324**, 163-166 (1986).
- Acha-Orbea, H. & McDevitt, H. O. *Proc. natn. Acad. Sci. U.S.A.* **84**, 2435-2439 (1987).
- Hardy, D. A., Bell, J. I., Long, E. O., Lindsten, T. & McDevitt, H. O. *Nature* **323**, 453-455 (1986).
- Schreuder, G. M. T., Doxiadis, I., Parlevliet, J. & Grosse-Wilde, H. in *Histocompatibility Testing* (eds Albert, E. D., Baur, M. P. & Mayr, W. R.) 243-248 (Springer, Berlin, 1984).
- Matsuki, K., Juji, T. & Satake, M. *Hum. Immunol.* **18**, 247-256 (1987).
- Maeda, H., Hirata, R., Thompson, A. & Mukai, R. *Hum. Immun.* **15**, 1-15 (1986).
- Bach, F. H., Ohta, N., Anichini, A. & Reinsmoen, N. L. in *HLA Class II Antigens* (eds Solheim, B. G., Moller, E. & Ferrone, S.) 249-265 (Springer, Berlin, 1986).
- Baur, M. P., Nagebauer, M. & Albert, E. D. in *Histocompatibility Testing* (eds Albert, E. D., Baur, M. P. & Mayr, W. R.) 677-760 (Springer, Berlin, 1984).
- Tanigaki, N., Tosi, R., Centis, D. & Ferrara, G. B. *Hum. Immun.* **12**, 47-57 (1985).
- Michelsen, B. & Lernmark, A. *J. clin. Invest.* **79**, 1144-1152 (1987).
- Cohen, N., Brautbar, C., Font, M.-P., Dausset, J. & Cohen, D. *Immunogenetics* **23**, 84-89 (1986).
- Moriuchi, J., Moriuchi, T. & Silver, J. *Proc. natn. Acad. Sci. U.S.A.* **82**, 3420-3424 (1985).
- Wicker, L. S. *et al. J. exp. Med.* **165**, 1639-1654 (1987).
- Estess, P., Begovich, A. B., Koo, M., Jones, P. P. & McDevitt, H. O. *Proc. natn. Acad. Sci. U.S.A.* **83**, 3594-3598 (1986).
- Gregersen, P. K. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 2642-2646 (1986).
- Larhammar, D. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 7313-7317 (1983).
- Long, E. O., Wake, C. T., Gorski, J. & Mach, B. *EMBO J.* **2**, 389-394 (1983).
- Todd, J. A. *et al. J. Bact.* **167**, 257-264 (1986).
- Spies, T., Sorrentino, R., Boss, J. M., Okada, K. & Strominger, J. L. *Proc. natn. Acad. Sci. U.S.A.* **82**, 5165-5169 (1985).
- Rollini, P., Mach, B. & Gorski, J. *Proc. natn. Acad. Sci. U.S.A.* **82**, 7197-7201 (1985).
- Bell, J. I. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 6234-6238 (1987).
- Auffray, C. *et al. Nature* **308**, 327-333 (1984).
- Okada, K. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 3410-3414 (1985).
- Tonnelle, C., DeMars, R. & Long, E. O. *EMBO J.* **4**, 2839-2847 (1985).
- Lee, B. S. M., Rust, N. A., McMichael, A. J. & McDevitt, H. O. *Proc. natn. Acad. Sci. U.S.A.* **84**, 4591-4595 (1987).
- Wu, S., Saunders, T. L. & Bach, F. H. *Nature* **324**, 676-679 (1986).
- Gorski, J. & Mach, B. *Nature* **322**, 67-70 (1986).
- Tieber, V. L., Abruzzini, L. F., Didier, D. K., Schwartz, B. D. & Rotwein, P. *J. biol. Chem.* **261**, 2738-2742 (1986).
- Freeman, S. M. *et al. Hum. Immun.* **19**, 1-6 (1987).
- Gubler, U. & Hoffman, B. J. *Gene* **25**, 263-269 (1983).
- Waye, M. M. Y., Verhoeyen, M. E., Jones, P. T. & Winter, G. *Nucleic Acids Res.* **13**, 8561-8571 (1985).
- Schenning, L. *et al. EMBO J.* **3**, 447-452 (1984).
- Lee, B. S. M., Bell, J. I., Rust, N. A. & McDevitt, H. O. *Immunogenetics* **26**, 85-91 (1987).
- Boss, J. M. & Strominger, J. L. *Proc. natn. Acad. Sci. U.S.A.* **81**, 5199-5203 (1984).
- Karr, R. W. *et al. J. Immunol.* **137**, 2886-2890 (1986).
- Turco, E. *et al. Immunogenetics* (in the press).

LETTERS TO NATURE

Alignment of radio and optical orientations in high-redshift radio galaxies

K. C. Chambers*, G. K. Miley†§ & W. van Breugel‡

* Department of Physics and Astronomy and Center for Astrophysical Sciences, The Johns Hopkins University, Baltimore, Maryland 21218, USA

† Space Telescope Science Institute, 3400 San Martin Drive, Baltimore, Maryland 21218, USA and Astrophysics Division, Space Sciences Department, European Space Agency, Noordwijk, The Netherlands

‡ Radio Astronomy Laboratory, University of California, Berkeley, California 94720, USA

The results of a VLA (Very Large Array) radio and CCD (charge-coupled device) optical imaging survey of ultra-steep spectrum radio sources confirm their association with faint and presumably distant galaxies. These galaxies have an almost universal alignment between the major axis of the optical emission and the radio axis. The alignment is far more significant than the known opposite

tendency for optically bright radio sources to be aligned with the minor axis of giant elliptical galaxies¹⁻³. Such a strong relationship between the radio emission and the large-scale optical properties of the galaxies suggests that high-redshift radio galaxies are fundamentally different from their low-redshift counterparts. This has important consequences for current ideas on the nature of high-redshift radio galaxies and the cosmological work based upon them.

The galaxies associated with powerful high-redshift radio sources have unusual morphologies—they are often elongated and even multimodal⁴. We have developed a technique for finding such objects by using the fact that ultra-steep-spectrum radio sources are systematically more luminous than normal steep-spectrum radio sources. The radio spectral index above 178 MHz correlates strongly with both redshift and radio luminosity. Furthermore, the fractional identification content (Sky Survey) and the linear size distribution are both dramatically smaller than for normal steep spectrum radio sources^{5,6}. We are carrying out a multispectral program to study the properties of radio sources selected for their ultra-steep radio spectra.

Our total sample consists of 51 objects, 33 from the 4C catalogue and 18 from the 3C catalogue, which have spectral indices $\alpha < -1$ between 178 MHz and 5,000 MHz. (Here α is defined by $S_\nu \propto \nu^\alpha$, where S_ν is the flux density at frequency ν .) They are confined in galactic longitude and latitude by $20^\circ < \delta < 80^\circ$ and

§ Permanent address: Sterrewacht, Postbus 9513, 2300 RA Leiden, The Netherlands.

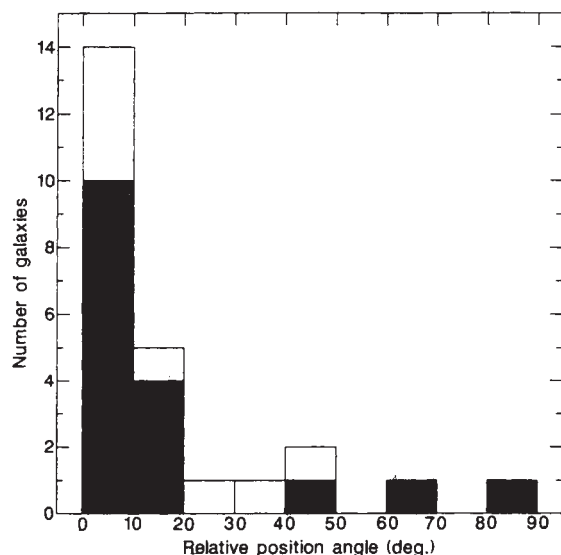


Fig. 1 Histogram of the angle between the optical and radio major axes for the sample of galaxies. The solid bars indicate those galaxies with well-defined position angles. The unshaded part includes galaxies whose position angle was obtained from the literature, and objects for which the error in relative position angle is greater than 10 degrees²⁸⁻³⁰.

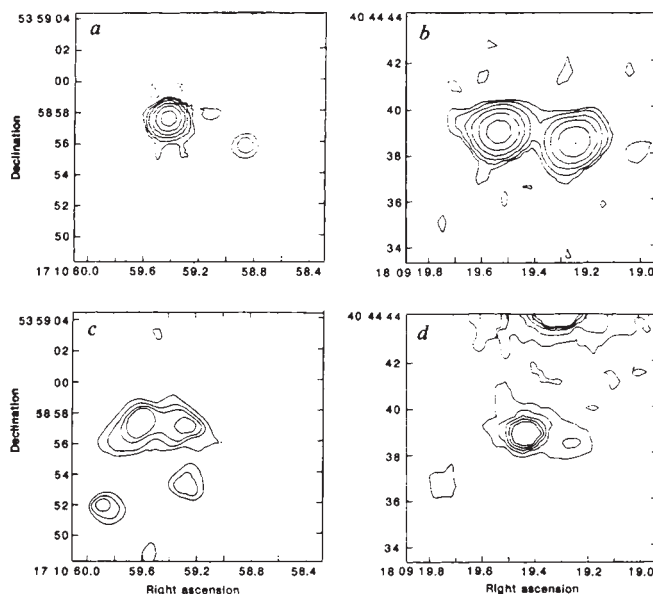


Fig. 2 Two examples of new identifications. *a*, A VLA radio image made at 4,885 MHz in the B array of 4C53.38. Contours are -2, 2, 4, 8, 16, 32, 64% of peak flux of 56 mJ. *b*, Isophotal contours from an R band CCD image of the galaxy. Contours are linearly spaced. Figure 2*c* and *d* are like *a* and *b* for galaxy 4C40.36, whose peak flux is 32 mJ. Vertical tick marks correspond to 2 arc s.

$|b''| > 15^\circ$. We have recently observed the entire sample at three wavelengths (20 cm, 6 cm and 2 cm), in matched arrays with the VLA. The initial results have borne out the assertion that almost all of the sources have the edge-brightened double morphologies (Cygnus A/Fanaroff-Riley class II⁷) characteristic of high luminosity radio sources, but there are also several bent and disturbed structures in our sample. Using these radio maps, we observed all 33 of the previously unidentified 4C sources with a CCD in the R band with the 2.1-m KPNO (Kitt Peak National Observatory) telescope in August 1986 and February 1987. Accurate astrometry has identified 19 extended objects with $19 < M_R < 23$. We do not yet have redshifts for our newly discovered objects but from their apparent magnitudes they are likely to have redshifts at least as high as the 3CR members of the sample, whose mean redshift is 1.01.

Using the subsample of 25 extended objects with $19 < M_R < 23$, we measured position angles of the extended emission within 3 arc s of the central maxima. To improve the signal to noise of the images, they were smoothed to 3 pixels (1.14 arc s). The position angle of the radio sources was measured between their outermost maxima. To minimize bias the optical position angles were measured independently by a third party. Figure 1 is a histogram of the difference between the radio and optical position angle. A Kolmogorov-Smirnov test shows the alignment in this histogram to be nonrandom at better than a 99.9% confidence level. Note that 12 of the 25 sources have a companion object within 10 arc s, which may aid in triggering the radio source⁸. The companion objects appear to have a random distribution of angles with respect to the radio axis, suggesting that confusion due to companions or close cluster members will add noise to the histogram. If some of our sources are confused by unresolved companions, our data are consistent with universal alignment between the projected optical and radio major axes. Figure 2 shows radio and optical contours of two examples.

This result is much more significant than, and opposite in direction to that found in studies of brighter and presumably lower-redshift radio galaxies in the Palomar Sky Survey. Careful studies convincingly show that the radio structures are preferentially but not invariably aligned along the projected minor axes in low-redshift elliptical galaxies¹⁻³, but this tendency may only

hold for those galaxies in which the compact radio cores coincident with their nuclei are prominent⁹. Even in normal elliptical galaxies it is difficult to determine the optical axis from ellipticity, since the isophote orientations may vary appreciably with radius¹⁰. Although the objects in our sample, are at the limit of detectability, their gross structure does not appear to resemble that of giant ellipticals. The unusual elongation in most cases leaves little doubt as to the direction of the major axis of the large-scale optical emission.

There is no evidence that the alignment we have found is specifically related to our selection criterion of ultra-steep radio spectra, rather than to the fact that this criterion is efficient at selecting intrinsically powerful radio sources. The fundamental questions are, what is the nature of the extended optical emission, and why does it coincide with the presumed energy pathway into the radio lobes? There are several possibilities. (1) The alignment is produced because the optical images include continuum emission from the synchrotron jets^{11,12}. Extrapolating the radio flux into the optical band with a spectral index of -0.6, the most optimistic case, gives *R* magnitudes greater than 20. Although it cannot be ruled out, the possibility that optical synchrotron jets are responsible for the alignment is unlikely except for a few cases. (2) The alignment is due to ionized gas is associated with the radio source. Morphological associations between extended radio and optical line emission regions have been seen in radio galaxies and in Seyfert galaxies¹³⁻²⁰. We have shown elsewhere that high luminosity radio sources are capable of entraining and accelerating significant amounts of material¹¹⁻¹⁶. (3) The alignment is due to continuum starlight from starforming regions associated with the radio sources. There is evidence that star formation in one gas-rich dwarf may have been triggered by the radio jet of a nearby elliptical^{21,22}. Similar jet-enhanced star formation, if more widespread at earlier epochs, could cause the alignment found in this survey. (4) The observed elongated morphologies are caused by dust lanes perpendicular to the radio axes.

Although only a small number of low redshift giant ellipticals with powerful radio sources show a band of absorption crossing the galaxy, a study of eight low-redshift radio galaxies with dust lanes shows that the seven unambiguous cases have dust lanes roughly perpendicular to the axes of the associated radio sources²³⁻²⁵. This is particularly relevant because of the apparent double appearance of the optical emission due to obscuration by a wide dust lane, as in the well-known powerful radio galaxy Cygnus A. If Cygnus A were at the distance of the galaxies in our sample, it would appear multimodal and slightly elongated in the direction of the presumed energy pathway into the radio lobes²⁶. Evidence that many powerful radio galaxies are indeed dusty is provided by the recent observations that they have a far infrared excess. (D. A. Golombek, G.K.M. and G. Neugebauer, in preparation.)

So far there are not sufficient data to distinguish decisively between these possibilities. Spectroscopic observations of objects in the sample are crucial, and are being carried out.

In conclusion, the discovery that the optical isophotes of distant powerful radio galaxies are aligned along the radio axis suggests that the optical and radio emission are intimately related. Considerable caution should therefore be exercised in treating high redshift radio galaxies as standard candles or as less evolved versions of giant elliptical galaxies. The existence of such alignments, the presence of metal absorption line systems with $z_{\text{abs}} \sim z_{\text{em}}$ and the bent QSOs may all be manifestations of what was at high redshift a more fertile circumgalactic environment associated with galaxy formation.

A similar alignment between radio and optical morphologies has been seen independently in high redshift 3CR galaxies by McCarthy, van Breugel, Spinrad, and Djorgovski (in preparation). Because our sample of 25 objects has only six 3CR radio galaxies, the occurrence of the alignment in both samples reinforces its significance.

We thank D. A. Golombek for independent measurement of the position angles and the staffs of the VLA and KPNO for their assistance in the observations. K.C.C. and G.K.M. are visiting astronomers at VLA and NRAO, and W.v.B. is a visiting astronomer at VLA. The National Radio Astronomy Observatory is operated by Associated Universities Inc and the National Optical Astronomy Observatory is operated by AURA, both under contract with the US National Science Foundation.

Received 3 June; accepted 24 August 1987.

- Palimaka, J. J., Bridle, A. H., Fomalont, E. B. & Brandie, G. W. *Astrophys. J.* **231**, L7-L11, (1979).
- Guthrie, B. N. G. *Mon. Not. R. astr. Soc.* **187**, 581-599 (1979).
- Shaver, P. et al. in *IAU Symp.* **97**, 55-57 (Reidel, Dordrecht, 1982).
- Spinrad, H. *Publ. astr. Soc. Pacif.* **98**, 269-284 (1986).
- Tielens, A. G., Miley, G. K. & Willis, A. G. *Astr. Astrophys. Suppl.* **35**, 153-162 (1979).
- Blumenthal, G. & Miley, G. K. *Astr. Astrophys.* **80**, 13-21 (1979).
- Fanaroff, B. L. & Riley, J. M. *Mon. Not. R. astr. Soc.* **167**, 31p-34p (1974).
- Heckman, T. et al. *Astrophys. J.* **311**, 526-547 (1986).
- Kapahi, V. K. & Saikia, D. J. *Astr. Astrophys.* **3**, 161-171 (1982).
- King, I. R. *Astrophys. J.* **222**, 1-13 (1978).
- Butcher, H. R., van Breugel, W. & Miley, G. K. *Astrophys. J.* **235**, 749-754 (1980).
- Miley, G. K. in *Astrophysical Jets* (eds Pacholczyk, A. G. & Ferrari, A.) 99-112 (Reidel, Dordrecht, 1983).
- Heckman, T. M., van Breugel, W., Miley, G. K., Balick, B. & Butcher, H. R. *Astrophys. J.* **262**, 529-553 (1982).
- van Breugel, W., Heckman, T. M. & Miley, G. K. *Astrophys. J.* **276**, 79-91 (1984).
- van Breugel, W., Heckman, T. M., Butcher, H. R. & Miley, G. K. *Astrophys. J.* **277**, 81-91 (1984).
- Heckman, T. M., van Breugel, W. & Miley, G. K. *Astrophys. J.* **286**, 509-516 (1984).
- van Breugel, W., Miley, G. K., Heckman, T. M., Butcher, H. R. & Bridle, A. H. *Astrophys. J.* **290**, 496-516 (1984).
- van Breugel, W., Heckman, T. M. & Miley, G. K. *Astrophys. J.* **311**, 58-84 (1986).
- van Breugel, W. *Can. J. Phys.* **64**, 392-397 (1986).
- Baldwin, J. A., Wilson, A. S. & Whittle, M. *Astrophys. J.* (in the press).
- van Breugel, W., Filippenko, T., Heckman, T. M. & Miley, G. K. *Astrophys. J.* **293**, 83-93 (1985).
- Brodie, J., Bowyer, S. & McCarthy, P. *Astrophys. J.* **293**, L59-L63 (1985).
- Ekers, R. D., Goss, W. M., Kotanyi, C. G. & Skellern, K. J. *Astr. Astrophys.* **69**, L21-L24 (1978).
- Kotanyi, C. G. & Ekers, R. D. *Astr. Astrophys.* **73**, L1-L3 (1979).
- Kotanyi, C. G. *Astr. Astrophys.* **74**, 156-160 (1979).
- Kronberg, P., van den Bergh, S. & Butten, S. *Astr. J.* **82**, 315-318 (1977).

Interstellar calcium towards supernova 1987A in the Large Magellanic Cloud

Pierre Magain

European Southern Observatory, La Silla, Casilla 19001, Santiago 19, Chile

We present high-resolution and high signal-to-noise spectra of interstellar calcium towards SN1987A in the Large Magellanic Cloud (LMC). They supplement previously published observations^{1,2} for Ca II and provide completely new results for the Ca I spectrum, in which eleven components are detected, not only at velocities corresponding to our Galaxy and the LMC, but also at intermediate velocities. Our spectra, that allow us to estimate the ionization balance in these interstellar clouds, provide some clues about their physical state and location. In particular, the components between 150 and 200 km s⁻¹ show a much lower ionization degree than other components. This may be due to recent compression of the gas by a shock, possibly associated with a former supernova explosion. This interpretation would require the corresponding clouds to be located inside the LMC, indicating that at least some intermediate velocity components are not of halo origin.

The observations were carried out at the European Southern Observatory (La Silla, Chile) on four nights between 16 March and 5 May 1987, with the Coudé Echelle Spectrometer fed by the 1.4-m Coudé Auxiliary Telescope. The detector was a high-resolution RCA charge-coupled device (CCD) (640×1,024 pixels, 15×15 µm each). Spectra were obtained around the H and K lines of Ca II (3,968 and 3,934 Å) and around the resonance line of Ca I (4,227 Å). The wavelength calibration was provided by a thorium lamp, with an accuracy of 0.2 km s⁻¹ internally and 1 km s⁻¹ externally. The resolving power, measured using the thorium lines, is 85,000 for the Ca II spectra, corresponding to 3.5 km s⁻¹ at full-width half-maximum (FWHM). For the much weaker Ca I line, the high signal needed, together with the less than optimal seeing conditions, dictated the choice of a slightly lower resolving power ($R = 65,000$, equivalent to 4.5 km s⁻¹). Several spectra were obtained for each spectral region, the total exposure time reaching 1 h, 1 h 30 min and 5 h 15 min for the H, K and Ca I lines, resulting in signal-to-noise ratios of 250, 200 and 550 respectively.

Figure 1 shows the spectra, in heliocentric velocity scale. The Ca I spectrum has been expanded 20 times to reveal the weak features. The measured velocities for Ca II agree to within 1 km s⁻¹ with the results of refs 1 and 2, except in some cases of severe blending, where no reliable velocity can be determined. We report, however, the detection of three additional weak components (Table 1), which can also be seen on the plots of ref. 1, but our most interesting results concern the Ca I spectrum, where 11 components are detected. Their velocities and equivalent widths, along with the corresponding quantities for Ca II, are listed in Table 2, which also gives the deduced column densities, ionization degree and electron density. Also included

Table 1 Additional Ca II components towards SN1987A

Ca II-H		Ca II-K	
$V_{\odot}$ (km s ⁻¹)	W (mÅ)	$V_{\odot}$ (km s ⁻¹)	W (mÅ)
81.3	7.2	80.9	15.6
86.3		86.6	
193.2	1.6	192.4	2.9

$V_{\odot}$ is the heliocentric velocity and W is the equivalent width.

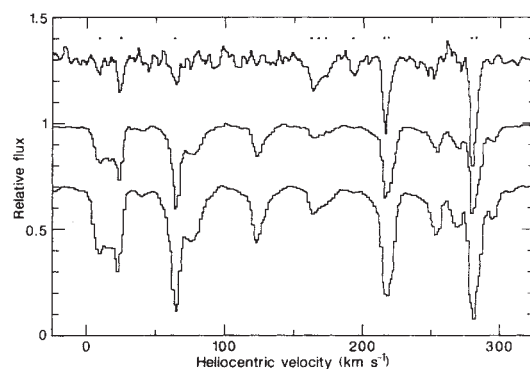
Table 2 Interstellar Ca I absorption towards SN1987A

Ca I		Ca II-H		Ca II-K		$N(\text{Ca I})$ $\times 10^{-8}$ (cm^{-2})	$N(\text{Ca II})$ $\times 10^{-11}$ (cm^{-2})	$N(\text{Ca I})$ $N(\text{Ca II})$	$n_e^\dagger$ ($\times 25/\beta$) (cm^{-3})
$V_\odot$ (km s^{-1})	W^* ($\text{m}\text{\AA}$)	$V_\odot$ (km s^{-1})	W ($\text{m}\text{\AA}$)	$V_\odot$ (km s^{-1})	W ($\text{m}\text{\AA}$)			$\times 10^3$	
9.7	0.3	8.3	16.9	8.0	31.4	9.0	4.0	2.4	0.059
24.9	0.6	23.6	18.7	23.5	28.7	22.0	5.6	3.8	0.095
64.3	0.5	64.8	33.7	64.5	59.8	19.0	8.3	2.3	0.057
—	<0.2	123.5	20.0	123.6	39.0	<7.0	4.4	<1.6	<0.04
163.6	1.2	164.1	12.8	163.4	26.2	43.0	2.8	15.0	0.38
168.6		170.3		169.9					
173.8		176.2		176.7					
193.8	0.3	193.2	1.6	192.4	2.9	12.0	0.3	36.0	0.91
216.0	1.4	216.3	41.2	216.9	67.2	51.0	11.0	4.5	0.11
218.9		221.1		221.8					
279.7		280.1		280.1					
283.1	2.1	285.2	49.2	286.0	81.2	76.0	13.0	5.7	0.14

* The r.m.s. uncertainties on the Ca I equivalent widths, as deduced from the measured signal-to-noise, amount to $\sim 0.1 \text{ m}\text{\AA}$. All reported detections are thus at least at the 3σ level (σ is standard deviation).

† Electron densities are estimated using $\beta = 25 \text{ cm}^{-3}$. This value, which is appropriate for the galactic clouds, might lead to some underestimate of n_e in shocked LMC gas.

Fig. 1 Spectra of Ca I (top), Ca II-H (middle) and Ca II-K (bottom) on a common heliocentric velocity scale. The ordinate scale refers to the middle spectrum, while the others are displaced vertically by 0.3 units. The Ca I spectrum is expanded 20 times. Dots mark the positions of the Ca I components listed in Table 2.



is a strong Ca II line with no Ca I counterpart detected. The velocities and equivalent widths were determined by fitting multiple gaussian profiles. For all Ca I and the weakest Ca II lines, the column density is directly proportional to the equivalent width. For the stronger Ca II lines, it has been estimated by the doublet ratio method³. Although this method may lead to significant errors⁴, this should not be the case here as the lines are not strongly saturated. The electron density was estimated using the formula $n_e = \beta N(\text{Ca I}) / N(\text{Ca II})$, with $\beta = 25 \text{ cm}^{-3}$ (ref. 5).

For the Ca II components with no Ca I counterpart detected, a typical upper limit on the electron density is 0.1 cm^{-3} . It is thus seen from Table 2 that all Milky Way (for which the heliocentric velocity $V_\odot < 25 \text{ km s}^{-1}$) and LMC ($V_\odot > 235 \text{ km s}^{-1}$) components have a rather low electron density (that is a high degree of ionization). This is also true for most of the intermediate velocity clouds, with the exception of those with radial velocities between $150\text{--}200 \text{ km s}^{-1}$, which show high electron densities. Such high densities have been observed in the line of sight to a few early-type stars⁵ and interpreted as the result of a shock wave, produced either by an expanding H II region or a supernova remnant. If this interpretation is correct, it is probable that the corresponding clouds are located in regions of recent star formation or, at least, in regions of non-negligible stellar density. Two additional points are worth mentioning. First, the similarity of the Ca I and Ca II profiles for the low ionization group around 170 km s^{-1} suggests that this group is not merely a blend from several distinct components, but rather the signature of a single high-density cloud, with a velocity dispersion of $\sim 15 \text{ km s}^{-1}$. On the other hand, from the survey of Prevot *et al.*⁶, no star within one degree of SN1987A is found

to have a radial velocity between $150\text{--}200 \text{ km s}^{-1}$. Thus, if located inside the LMC, the high-density clouds have a high velocity with respect to the neighbouring stars. These two facts fit nicely into the supernova remnant hypothesis. On the basis of the available evidence, we can thus conclude that at least some of the intermediate velocity clouds are not located in the halo, but belong to the LMC, in agreement with the results of a recent Ca II survey⁷.

Finally, it is worth noting that this hypothetical shock might be associated with the 'superbubble' N157C = 0536–692^{8,9} which is centred at <6 arc min from SN1987A and has a radio diameter of the same order. This 'superbubble' is believed to be the result of one (or several) supernovae, and of the winds of the early-type stars from the OB association NGC2044 located inside it. Although the radio emission does not seem to extend to the position of SN1987A, the optical emission does, as is revealed by H α and [O III] exposures^{9,10}. If one, or several, of the high-density clouds is associated with this nebulosity, this would imply that SN1987A is located behind (or inside) N157C.

We thank M. Heydari-Malayeri and H. Lindgren for stimulating discussions and critical comments.

Received 1 June; accepted 29 July 1987.

1. Andreani, P., Ferlet, R. & Vidal-Madjar, A. *Nature* **326**, 770–772 (1987).
2. Vidal-Madjar, A. *et al. Astr. Astrophys.* **177**, L17–L20 (1977).
3. Spitzer, L. *Physical Processes in the Interstellar Medium* (Wiley, New York, 1978).
4. Nachmann, P. & Hobbs, L. M. *Astrophys. J.* **182**, 481–487 (1973).
5. Federmann, S. R. & Hobbs, L. M. *Astrophys. J.* **265**, 813–816 (1983).
6. Prevot, L. *et al. Astr. Astrophys. Suppl.* **62**, 23–37 (1985).
7. Songaila, A., Blades, J. C., Hu, E. M. & Cowie, L. L. *Astrophys. J.* **303**, 198–215 (1986).
8. Mathewson, D. S. *et al. Astrophys. J. Suppl. Ser.* **58**, 197–200 (1985).
9. Lortet, M. C. & Testor, G. *Astr. Astrophys.* **139**, 330–340 (1984).
10. Elliott, K. H., Goudis, C., Meaburn, J. & Tebbutt, N. J. *Astr. Astrophys.* **55**, 187 (1977).

Speckle interferometric observations of supernova 1987A and of a bright associated source

W. P. S. Meikle, S. J. Matcher & B. L. Morgan

Astrophysics Group, Blackett Laboratory,
Imperial College of Science and Technology, Prince Consort Road,
London SW7 2BZ, UK

The explosion of SN1987A in the Large Magellanic Cloud¹ provides a unique opportunity for the direct measurement, by means of interferometry, of the spatial distribution and evolution of the optical emission of a supernova. In particular, such observations may help to resolve the controversy concerning the physics of the hydrogen emission-line region^{2,3}. Also, as the supernova fades there is the prospect of obtaining the intensity structure of any light echo arising from dust-scattered radiation⁴. Here we report optical speckle interferometric observations of SN1987A on days 38, 47 and 50 after the explosion on 23 February 1987. The supernova itself was unresolved, with angular radius upper limits of ~ 0.012 arc s at 3,869 Å, 0.015 arc s at 4,861 Å and 4,921 Å, and 0.020 arc s at 5,875 Å and 6,585 Å, consistent with angular radii derived from photometry and spectroscopy. On day 30 after the explosion a group from the Harvard-Smithsonian Center for Astrophysics (CfA) discovered a second source lying at a small angular displacement from the supernova^{5,6}. They again observed it on day 38. On day 50 we observed a second source at 6,585 Å. We measure its offset from the supernova to have been 0.074 ± 0.008 arc s and estimate that it was ~ 3 mag fainter than the supernova at 6,585 Å at the same epoch. Such a bright source was not present before the explosion, implying that it was caused by the supernova. The source we observed was similar in magnitude and position angle to the feature discovered by the CfA group, although we find the angular displacement to have been 0.015 ± 0.011 arc s larger.

The observations were made with the Imperial College Speckle Interferometer^{7,8} attached to the Anglo-Australian Telescope (AAT) during April 1987. A summary of the observation parameters is given in Table 1. A comparison star was measured immediately before and after every supernova observation. Particular care was taken to minimize the effect of atmospheric dispersion. Through the use of non-deviating but dispersing prisms and narrow passbands (see Table 1), and by making all observations above an altitude of 34° , elongation of the image due to atmospheric dispersion was always substantially less than the size of an AAT diffraction-limited image. The data were recorded on video cassettes and subsequently processed using our dedicated hard-wired autocorrelator. The two-dimensional autocorrelation function (AF) for each image is generated by measuring the magnitude and direction of the 'vector' from every pixel containing a photon event to every other pixel containing an event within the frame. The AFs of all the frames for an integration at a given wavelength are then summed.

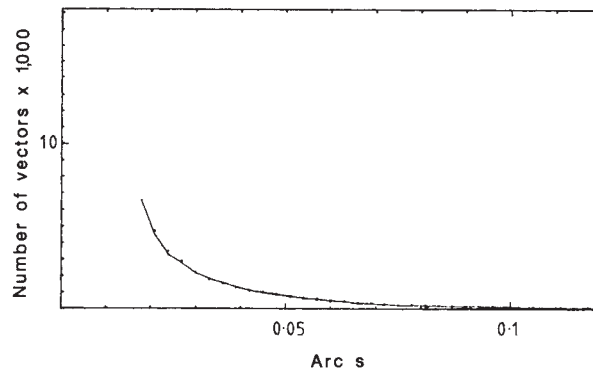


Fig. 1 Averaged autocorrelation functions of SN1987A (dots) and an unresolved reference star, BS2015 (solid line) after subtraction of a near-gaussian seeing component. Averaging was carried out between position angles $300^\circ/120^\circ$ and $60^\circ/240^\circ$. The wavelength was 4,921 Å. The close similarity between the two functions demonstrates the lack of any resolved disk or ring in the supernova image. For angular displacements smaller than about 0.015 arc s, the AFs are not shown because they are dominated by the photon spike (see ref. 8).

One-dimensional profiles at specific position angles (PAs) can be selected from these AFs. (In this data reduction procedure there is a 180° ambiguity in the selected PA values).

To test for the presence of a resolved disk or ring, one-dimensional profiles through the two-dimensional AFs were azimuthally averaged at each wavelength over PAs between $300^\circ/120^\circ$ and $60^\circ/240^\circ$. The seeing component of the AF was modelled by a function of the form $\exp(-r^{5/3})$ (ref. 8), which was fitted to the observed AF at large angular offset (0.4–0.5 arc s). Subtraction of this model seeing disk from the AF leaves a residual which represents the AF of the object intensity distribution. A typical residual AF is shown in Fig. 1 together with that for the comparison star BS2015. The similarity of the forms of the residual AFs demonstrates the absence of any resolved disk or ring down to an angular radius about equal to that of the Airy disk for the AAT. Upper limits for the angular radius of SN1987A are shown in Table 2. (Some of these limits have already been reported in ref. 9). Also shown in Table 2 are angular radii determined both from VRIJKL photometry¹⁰, assuming the supernova spectrum to be essentially black body and from the blue-shifts of the H α -P-Cygni trough¹¹, assuming constant velocity since the explosion. Clearly our upper limits are consistent with the indirectly determined angular radii. The photospheric radius of SN1987A was exceptionally small. This is believed to have been due to the compact nature of the blue supergiant progenitor^{12,13}. Expansion from this compact configuration resulted in greater adiabatic cooling than would occur in the more extended red supergiant usually held to be the progenitor of type II events. Consequently the photosphere lay much deeper in the ejecta than it would for a more typical type II at a similar epoch. Catchpole *et al.* (preprint, South African Astronomical Observatory (SAAO), 1987) report that

Table 1 Observation log

Date	Epoch (days)	Approx. time (UT)	Wavelength (Å)	Bandwidth (Å)	Frame exposure time (ms)	Integration time (min)
2 April	38	10.00	3,869*	15	40	12
		11.20	4,861	16	40	13
11 April	47	09.55	4,861	16	10	10
		10.30	4,921	15	10	10
14 April	50	10.05	5,876	10	40	8
		10.50	6,585	10	40	7

* Reference stars were SA0249316 and BS2015; at other wavelengths reference star was BS2015.

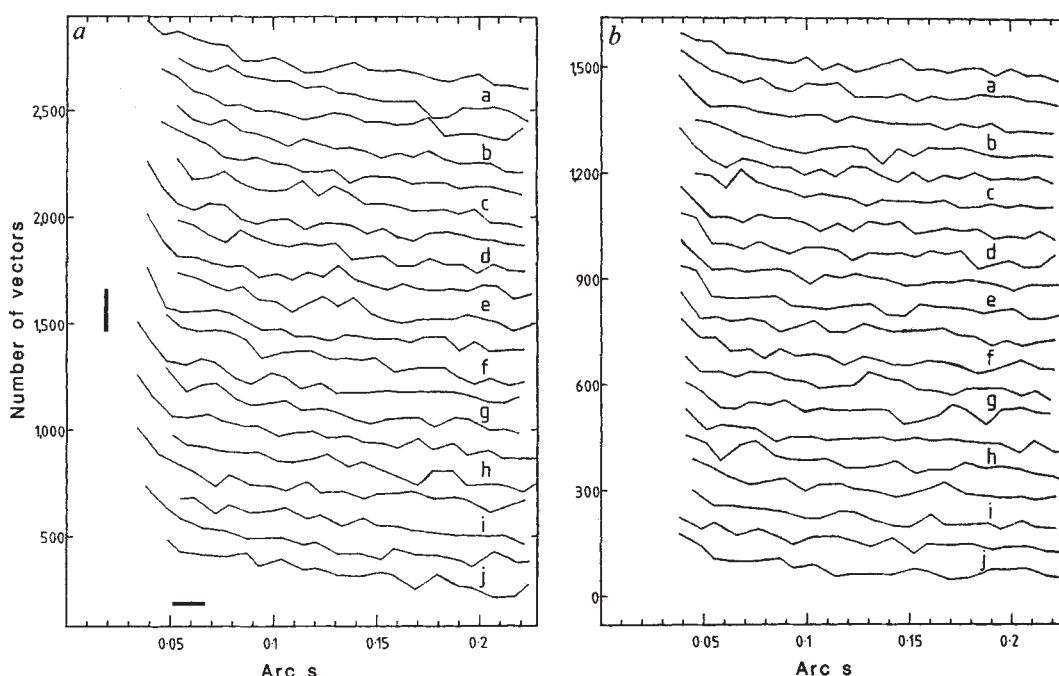
Table 2 Upper limits for the angular radius of SN1987A

Date	Epoch (days)	Wavelength (Å)	Angular radius (speckle) (arc s)	Angular radius (photometry)* (arc s)	Angular radius (H α)† (arc s)
April 2	38	3,869	<0.012	0.0012	0.0030
		4,861	<0.015		
April 11	47	4,861	<0.015	0.0014	0.0035
		4,921	<0.015		
April 14	50	5,876	<0.020	0.0014	0.0036
		6,585	<0.020		

* From VRIJHL photometry assuming black-body emission¹⁰.

† From the blue-shifts of the H α absorption trough¹¹ assuming constant velocity since the explosion of the supernova.

Fig. 2 Profiles through two-dimensional autocorrelation functions of SN1987A at 6,585 Å (*a*) and of reference star BS2015 at 5,876 Å (*b*). Both sets of data were obtained on 14 April 1987. The letters a-j on the diagrams each indicate a pair of profiles having the same position angle, but arising from odd (above letter) or even (below letter) rasters of the video frame. The position angles range from 351°/171° (*a*) to 36°/216° (*j*), at 5° intervals. In each figure the 'number of vectors' scale applies to the upper j AF, and for clarity the other AFs have been displaced with respect to it. For angular displacements smaller than about 0.035 arc s the AFs are not shown because they are dominated by the photon spike (see ref. 8). In *a*, a zone showing the signature of a secondary source lies between 0.06 and 0.09 arc s and between 11°/191° (*e*) and 26°/206° (*h*). The bars lying alongside the axes in *a* indicate the location of the source discovered by the CfA group⁶. The lengths of the bars represent their quoted uncertainties in angular displacement and position angle.



the photosphere reached a maximum angular radius of 0.0017 arc s about 100 days after the explosion. This is much smaller than the Airy disk of a 4-m telescope, but it is quite possible that the emission-line regions, especially those of hydrogen, will exceed the Airy disk size within the next year.

On 4 May 1987 Karovska *et al.* (CfA), reported their discovery⁵ of a bright feature at a small angular displacement from the supernova position. Their initial data reduction revealed detections at wavelengths of 6,563 Å and 5,330 Å. At that time we had not yet begun to reduce our data in this region of the spectrum. After the CfA report, we began analysis of the corresponding wavelength region of our observations and in Fig. 2 we show some preliminary results. We stress that these results represent only an initial phase of the full data reduction, which will take some time to complete, but we consider that the importance of these observations justifies this presentation of partially reduced results.

Figure 2 displays a representative sample of AFs obtained from our observations of the supernova (Fig. 2*a*) and reference star BS2015 (Fig. 2*b*). The AFs are arranged in pairs in order of PA, the respective members of each pair resulting from the odd and even rasters of the video frame at a given PA. In each

figure, the PA interval between each pair of AFs is 5°, and the total PA range is 351°/171° to 36°/216°. Fig. 2*a* shows the AFs obtained from our 6,585 Å supernova observations made on 14 April 1987. Given the size of the AAT Airy disk and our selected PA intervals, an unresolved secondary source of small magnitude difference and high signal-to-noise ratio would be present in Fig. 2*a* as a 'hump' in the AFs, extending over about 0.04 arc s and persisting over about 6 odd and 6 even AFs lying adjacent to each other. But when the signal-to-noise ratio is as low as it is for our 6,585 Å data, and for a secondary source of the large magnitude difference seen by the CfA group, we would expect the source to be apparent over a smaller range of angular offsets and PAs. Noise-induced fluctuations in the AFs should generally not correlate between adjacent pixels (pixel width = 0.008 arc s) nor between AFs 5° or more apart, nor between odd and even AFs at the same PA.

Inspection of Fig. 2*a* reveals a zone within which the AFs exhibit a hump, correlated over several odd and even AFs, and located between angular displacements of 0.06 arc s and 0.09 arc s and between PAs of 11°/191° and 26°/206°. We interpret this as evidence of a secondary source about 3 magnitudes fainter than the supernova. We estimate that the source has a

PA of $16^\circ/196^\circ \pm 2^\circ$ and, following correction for the atmospheric seeing contribution to the AF (see above), an angular offset of 0.074 ± 0.008 arc s from the supernova. (In ref. 14 the reported angular displacement is wrong owing to a mistake made in applying the seeing correction; see ref. 15). In Fig. 2a we have also indicated the CfA source position by means of the two bars along the axes. (The lengths of the bars represent their quoted uncertainties in angular displacement and PA.) Within the errors, the angular displacement and PA of our candidate secondary source agree with those reported by the CfA group^{5,6}. The total PA range we searched was $296^\circ/116^\circ$ to $56^\circ/236^\circ$ (more than twice that shown in Fig. 2) but no other candidate sources were revealed in either the supernova or reference star AFs. The failure to find a zone in the reference star data showing the signature of a secondary source is evidence that the correlation zone in Fig. 2a is not an instrumental effect such as a filter second reflection. The reference star data were obtained about one hour before the supernova data under identical instrumental conditions, except that a 5,876 Å filter was used. The reference star data at 6,585 Å turned out to be too noisy to be usable, partly due to poor observing conditions. Nevertheless, we consider it extremely unlikely that the correlation zone was due to a second reflection in the 6,585-Å filter because, in the Imperial College Speckle Interferometer, the beam which passes through the filter is collimated. Moreover, in other studies using the Imperial College system no second reflection effect has ever been observed.

We believe that the data presented in Fig. 2 support the reality of the secondary source discovered by the CfA group. Henceforth we shall refer to the secondary source as the 'companion', although this does not necessarily imply close physical proximity to the supernova as the binary signature could simply be a line-of-sight effect. (We also note that the 180° ambiguity in the PAs means that the correlation zone in Fig. 2a is also consistent with two secondary sources lying symmetrically on opposite sides of the supernova). On April 13.75 the supernova magnitude in the R-band was 2.74 (ref. 10), and given that the 6,585 Å filter lay close to the H α emission peak, the companion was probably brighter than 6 mag. We find no evidence that either the supernova or the companion is resolved, implying angular radii of no more than ~ 0.02 arc s. Our supernova observations at 3,869 Å, 4,861 Å, 4,921 Å and 5,876 Å have also been given a

wavelengths (6,563 Å and 6,585 Å, respectively) and with similar magnitudes. The second of our possible PA values ($196^\circ \pm 2^\circ$) is consistent with their value. The angular displacement which we observe on day 50 after the explosion is 0.015 ± 0.011 arc s larger than the value they obtained on days 30 and 38 at a similar wavelength. The CfA group also observed a companion source at 4,500 Å and 5,330 Å, at about the same position but with a greater difference in magnitude relative to the supernova than at 6,563 Å. This is not in conflict with our non-detection of the companion at other wavelengths because the signal-to-noise ratio of the CfA results⁶ is higher than that of our results at this stage of our data processing.

A successful model for the companion to SN1987A must account for the properties now described. The companion is very bright. At 6,585 Å it has a magnitude of ≤ 6 , ~ 100 times brighter than any star within at least an arc min of the supernova position before the explosion. We therefore conclude that the companion was caused by SN1987A. The angular displacement which we observe only 50 days after the explosion requires the cause to have moved faster than $0.4c$. The companion is clearly not the result of normal supernova ejecta. Velocities only up to $31,000 \text{ km s}^{-1}$ have been observed in SN1987A (ref. 11).

From our observations we can place constraints on the location, motion and linear size of the companion on 14 April 1987. Taking the supernova-Earth distance to be 50 kpc¹⁶, the observed angular displacement of 0.074 ± 0.008 arc s means that the companion could not have been closer than 21.5 ± 2.3 light days from the supernova. If it lay further from the Earth than the supernova then it could not have been more than 29.6 ± 1.0 light days from the supernova. A greater distance would mean that the delay in the appearance of the companion after the earliest supernova observation would have exceeded 50 days. If the companion is nearer to the Earth than the supernova, then light travel time arguments do not constrain the companion's location. But as we bring the companion away from the supernova, the amount of flux it can intercept from the supernova falls off as r^{-2} which presents increasing difficulties for models in which the companion re-radiates the supernova flux. In addition the companion must be at least a few parsecs from the Earth otherwise parallactic movement between days 30 and 50 would probably have produced a much greater change in position relative to the supernova than was observed.

If we interpret the different angular displacements observed by the two groups as being due to movement which began at the supernova when it exploded, then a least-squares straight line fit gives a gradient of 0.00159 ± 0.00011 arc s per day, which corresponds to a transverse velocity of $(0.46 \pm 0.03)c$ at 50 kpc. But on the basis of the observations made between days 30 and 50 alone we cannot convincingly rule out the possibility that the angular displacement from the supernova was unchanged between these epochs. Because the companion is unresolved and no more than 50 kpc from Earth, its diameter perpendicular to the line of sight must be less than about 3×10^{16} cm.

If we assume the companion to be spherical, then from above the maximum solid angle it can subtend at the supernova is $\sim 2\%$ of 4π . Yet at 6,585 Å the flux from the companion is about 10% of the supernova flux. Thus it is difficult to account for the companion as a simple reflection of the supernova radiation from dust without invoking a rather contrived cloud geometry or non-isotropic emission from either object. However, a cloud of material might reprocess radiation from another part of the spectrum into the observed wavebands. A particularly intriguing possibility is that the strong extreme ultraviolet (EUV) pulse^{17,18} released during the first few hours of the explosion might heat and ionize a nearby gas cloud, which then re-radiates much of the absorbed energy into the optical region. However, the detection of the companion at 4,500 Å and 5,330 Å in 100-Å wide filters⁶ tends to argue against an 'H II-region-like' emission line spectrum. This is consistent with the lack of strong narrow emission lines in contemporary spectra¹⁰. Moreover, it has been

Table 3 Imperial College observations of the companion of SN1987A

Epoch (days)	Wavelength (Å)	Δm^*	Ang. displ. from SN1987A (arc s)	Position angle (degrees)
38	3,869	$> +4$	-	-
38, 47	4,861	$> +4$	-	-
47	4,921	$> +4$	-	-
50	5,876	$> +4$	-	-
50	6,585	$\sim +3$	0.074 ± 0.008	$16/196 \pm 2$

* $\Delta m = (\text{magnitude of companion}) - (\text{magnitude of supernova})$.

preliminary examination over the same range of PAs (including $16^\circ/196^\circ$) as was covered at 6,585 Å, but in spite of a better signal-to-noise ratio than at 6,585 Å, no correlation zone was found satisfying the criteria described above for binary signature. A summary of the observations is given in Table 3.

The companion source observed by the CfA group at 6,563 Å on 25 March and 2 April 1987 was 2.7 ± 0.2 magnitudes fainter than the supernova, and had an angular displacement of 0.059 ± 0.008 arc s and PA of $194^\circ \pm 2^\circ$ (ref. 6). Ambiguity in the PA was removed by their use of the Knox-Thomson image reconstruction technique. Thus the companion sources observed by the CfA and Imperial College groups were detected at similar

calculated¹³ that the small radius of the progenitor restricts the energy of the EUV flash to $\sim 10^{46}$ ergs, which is only $\sim 0.1\%$ of what is needed in the re-radiation model.

Further discussion of the nature of the companion is beyond the scope of this paper. Later speckle interferometric measurements have already been carried out by several groups and more are planned. The information that these observations will provide on the brightness evolution, spectrum and movement of the companion will surely clarify our understanding of this remarkable phenomenon.

We thank Russell Cannon, Peter Gillingham and the staff of the Anglo-Australian Observatory for their assistance and support in making these observations at very short notice. We also thank Russell Cannon, David Branch, Bob Kirshner, Peter Nisenson, Bob Noyes, John Quenby and Martin Rees for helpful discussions and David Allen, Phil Andrews, Andy Fabian, James Graham, Andy Longmore, Ken Reay, Jason Spyromilio and Peredur Williams for their encouragement and support. S.J.M. is grateful for support by an SERC research studentship. This work was supported by an SERC research grant.

Received 6 July; accepted 3 September 1987.

1. Shelton, I., Duhalde, O. & Jones, A. *IAU Circ. No.* 4316 (1987).
2. Branch, D. *et al. Astrophys. J.* **244**, 780–804 (1981).
3. Uomoto, A. & Kirshner, R. P. *Astrophys. J.* **308**, 685–690 (1986).
4. Chevalier, R. A. *Astrophys. J.* **308**, 225–231 (1986).
5. Karovska, M., Nisenson, P., Noyes, R. & Papaliolios, C. *IAU Circ. No.* 4382 (1987).
6. Nisenson, P., Papaliolios, C., Karovska, M. & Noyes, R. *Astrophys. J.* **320**, L15–L18 (1987).
7. Hebdén, J. C., Morgan, B. L. & Vine, H. A. *Proc. Society of Photo-Optical Instrumentation Engineers* **445**, 477–483 (1984).
8. Hebdén, J. C. thesis, Imperial College (1984).
9. Matcher, S. J., Meikle, W. P. S. & Morgan, B. L. *IAU Circ. No.* 4369 (1987).
10. Menzies, J. W. *et al. Mon. Not. R. astr. Soc.* **227**, 39P–49P (1987).
11. Dachs, J. & Hanuschik, R. *IAU Circ. No.* 4367 (1987).
12. Arnett, W. D. *Astrophys. J.* **319**, 136–142 (1987).
13. Woosley, S. E., Pinto, P. A. & Ensman, L. *Astrophys. J.* (in the press).
14. Matcher, S. J., Meikle, W. P. S. & Morgan, B. L. *IAU Circ. No.* 4391 (1987); erratum *IAU Circ. No.* 4394 (1987).
15. Matcher, S. J., Meikle, W. P. S. & Morgan, B. L. *IAU Circ. No.* 4413 (1987); erratum *IAU Circ. No.* 4414 (1987).
16. Rowan-Robinson, M. *The Cosmological Distance Ladder* (Freeman, New York, 1985).
17. Klein, R. I. & Chevalier, R. A. *Astrophys. J.* **223**, L109–L112 (1978).
18. Falk, S. W., *Astrophys. J.* **226**, L133–L136 (1978).

Circumstellar interaction and a pulsar nebula in the supernova 1986j

Roger A. Chevalier

Department of Astronomy, University of Virginia, PO Box 3818, Charlottesville, Virginia 22903, USA

The supernova SN1986j was first discovered as a radio source in August 1986¹, although there were a number of pre-discovery observations. Here I show that the circumstellar interaction model for radio supernovae provides an adequate fit to the existing data, implies a time of explosion of about January 1983, and provides a guide for the future evolution of the radio source. A shell-like source structure is expected. The model together with very-long-baseline interferometry (VLBI) observations of the radio source, implies a shock velocity of $\sim 10^4$ km s⁻¹ and a pre-supernova mass loss rate of $\sim 10^{-4} M_{\odot}$ yr⁻¹ for a wind velocity of 10 km s⁻¹. A search for an infrared dust echo is warranted and an X-ray luminosity of $\sim 10^{40}$ erg s⁻¹ is expected. The time behaviour of the optical flux and its high luminosity suggest that a central pulsar nebula is the energy source. The required properties are comparable to those expected in the early phases of the Crab nebula. The radiated energy from the pulsar nebula is absorbed by expanding, helium-rich material extending out to a velocity of ~ 700 km s⁻¹ or a radius of 8×10^{15} cm on September 1986. The radius of the pulsar nebula at that time is $\sim 5 \times 10^{15}$ cm.

The supernova 1986j in NGC891 is of special interest because it is currently the brightest extragalactic radio supernova and it

has a peculiar optical spectrum¹. The circumstellar interaction model has been successful in fitting the radio observations of other supernovae^{2–4}, and it can also be applied to this case. Chevalier³ presented model radio light curves based on the assumptions of a power-law supernova density profile, a constant ratio of magnetic and of relativistic electron energy density to thermal energy density, and a constant electron spectral index. Figure 1 shows a fit to the radio observations (listed in ref. 1) with an electron energy spectrum $N(E) \propto E^{-2.8}$ and shock radius expanding with time as $t^{1.0}$. This is the same model that gave the best fit to radio observations of SN1979c. The model fit has the supernova occurring in January 1983 and optical depth unity to circumstellar free-free absorption at an age, t_{20} , of 1,600 days. Fits with the explosion occurring in July 1982 or July 1983 are significantly worse. In this model, the radio spectrum at one time is expected to be of the form $x^{-0.9} \exp(-x^{-2})$, where $x = \nu/\nu_c$ and ν_c is the frequency at which the free-free optical depth is unity at the time of observation. This spectrum gives an excellent fit to observations¹ in August and September 1986 with $\nu_c = 1.95 \pm 0.05$ GHz. Hamilton *et al.*⁵ have also suggested that the radio emission is from circumstellar interaction, but they did not model the radio data.

The SN1986j radio source was resolved by VLBI techniques on 29 September 1986⁶. At 18 cm, the angular size was 0.0014 ± 0.0002 arc s (full width at half-maximum of a gaussian). If the distance to NGC891, D , is 0.56 times the distance to the Virgo Cluster (ref. 7) and the distance to the Virgo Cluster is 21 Mpc (ref. 8), then $D = 12$ Mpc. The measured radio radius of SN1986j is then 1.3×10^{17} cm. This value may be reduced because of the effects of interstellar scattering. The estimated radius must be modified if the emitting region is in a shell, as in the present model. The radius is increased by a factor of 1.8 if the emitting region is a uniform sphere (N. Bartel, personal communication) and it is then reduced by a factor of 1.3 for shell emission⁹, leading to a radius of 1.7×10^{17} cm. At an age of 1,360 days, the average velocity of the radio shell is 15,000 km s⁻¹. Allowing for some reduction in the actual radius due to interstellar scattering, the velocity is comparable to that inferred for SN1979c². If the velocities are comparable, then the wind density surrounding SN1986j is a factor of two higher than that surrounding SN1979c because of the higher value of t_{20} for SN1986j: 1,600 days versus 950 days. The high luminosity of SN1986j supports this interpretation. In the model described above, the radio luminosity at a specified age is proportional to $R^3 A^{2.9}$, where R is the radius of the radio shell and A is proportional to the circumstellar density. We estimate that in the optically thin regime, the radio luminosity of SN1986j is a factor of six higher than that of SN 1979c at the same age. This difference is well accounted for by the difference in circumstellar density. If the velocities in SN1986j were substantially lower than those in SN1979c, the radio luminosity would require a higher value of A for SN1986j, but the free-free absorption turn-on would require a lower value of A . We conclude that the wind density in SN1986j is about twice that in SN1979c and corresponds to a pre-supernova mass loss rate of $10^{-4} M_{\odot}$ yr⁻¹ for a wind velocity of 10 km s⁻¹.

SN1979c was a source of infrared emission, presumably from hot dust; the data can be adequately interpreted as an echo in a dusty circumstellar medium¹⁰. The model for SN1979c implied that the dusty envelope had an outer radius of 8×10^{17} cm, which results in a cutoff of the infrared luminosity at an age of 1.7 yr. For similar infrared emission to occur from SN1986j at an age > 4 yr, a dusty envelope extending beyond 2×10^{18} cm is required. This is possible, and a search for infrared emission may be warranted. At an age of 4 yr, a luminosity of $\sim 10^{40}$ erg s⁻¹ with a spectral peak near 10 μ m is possible. The circumstellar interaction is also expected to give rise to X-ray emission from shock-heated gas. From the models of ref. 11, we estimate an X-ray luminosity of $\sim 10^{40}$ erg s⁻¹ and a gas temperature of $(1-4) \times 10^7$ K. There is a factor of ≥ 3 uncertainty in the luminos-

ity due to uncertainty in the density profile of the supernova ejecta.

The optical data from SN1986j¹ are of special importance because spectroscopy of a supernova at an age of ~ 4 yr is unprecedented. The large observed $H\alpha/H\beta$ ratio suggests that the emitting gas has a high density, $n_e \geq 10^9 \text{ cm}^{-3}$. This, together with the low velocities of the matter ($< 1,000 \text{ km s}^{-1}$), indicates that the emission is from a central region of the supernova. Plausible energy sources for the radiation are radioactivity and a central pulsar. On 6 September 1986, the $H\alpha$ luminosity was $3 \times 10^{39} \text{ erg s}^{-1}$ if $A_v = 2$ mag (ref. 1), where A_v is the visual extinction in magnitudes. The total power available from the decay of ^{56}Co is $1.4 \times 10^{43} (M_{\text{Ni}}/M_\odot) \exp(-t/114 \text{ days}) \text{ erg s}^{-1}$, where M_{Ni} is the initial mass of ^{56}Ni synthesized in the explosion and t is the age. To produce the above $H\alpha$ luminosity at an age of 1,340 days, $M_{\text{Ni}} = 27 M_\odot$ is required. This is much larger than is expected in any stellar model. Also the r -band magnitude, which includes the $H\alpha$ line, changed from 18.4 on the night of 6–7 January 1984 to 19.5 on the night of 2–3 September 1986¹. This corresponds to an e-folding decay time of $\sim 1,000$ days, much longer than the decay time of ^{56}Co .

The high late luminosity and slow decay are suggestive of energy input from a central pulsar. A pulsar with an initial energy output of $L_i = 10^{40} \text{ erg s}^{-1}$ is expected to maintain this energy output for ~ 10 yr. The observed $H\alpha$ line profile shows that the $H\alpha$ emissivity drops with radius, but we approximate the gas density as being constant out to a velocity of $v_m = 700 \text{ km s}^{-1}$. At an age of 1,340 days, the radius of this inner region is $8 \times 10^{15} \text{ cm}$. The radius of a pulsar bubble containing relativistic fluid is then (ref. 12)

$$R_p = 5 \times 10^{15} \left(\frac{L_i}{10^{40} \text{ erg s}^{-1}} \right)^{1/5} \left(\frac{V_m}{700 \text{ km s}^{-1}} \right)^{3/5} \\ \times \left(\frac{M_c}{2M_\odot} \right)^{-1/5} \left(\frac{t}{1,340 \text{ days}} \right)^{6/5} \text{ cm},$$

where M_c is the mass of the constant-density core, if radiative losses from the relativistic fluid are neglected. Beyond R_p , the gas is in free expansion. The matter that is swept up by the pulsar bubble is subject to the Rayleigh–Taylor instability and may form filaments. The density of the uniform gas is $\sim 1.8 \times 10^{-15} \text{ g cm}^{-3}$ for $M_c = 2M_\odot$. The core gas is optically thick at ultraviolet and X-ray wavelengths, where much of the radiation from the pulsar bubble may be emitted. The initial radiation may be either from relativistic gas surrounding the pulsar or from the shock front driven into the slow-moving supernova gas. Detailed modelling of the pulsar nebula will be difficult because of a lack of knowledge of the density distribution and of the incoming radiation spectrum. The fact that the helium lines are more prominent than lines of other heavy elements¹ is unusual for supernovae and may imply a relatively large helium abundance. This is suggestive of an initial stellar mass in the range 8–13 $M_\odot$, as has been suggested for the Crab nebula¹³. The model described here for SN1986j strongly resembles the model for SN1054 and the Crab Nebula described in ref. 12.

Not all type II supernovae behave at late times like SN1986j. SN1980k in NGC6946 had an $H\alpha$ luminosity of $1 \times 10^{38} \text{ erg s}^{-1}$ at an age of 660 days¹⁴, for a distance of 10 Mpc and $A_v = 1.1$ mag. The rate of change of the $H\alpha$ flux in this case is consistent with the ^{56}Co decay time and radioactivity is the probable energy source. The low luminosity of SN1980k implies that either an energetic pulsar is not present, or the pulsar energy output is not coupled to the $H\alpha$ line emission.

The circumstellar interaction model for SN1986j does make clear predictions of the radio, infrared and X-ray behaviour of this object. The pulsar nebula model is more difficult to test. A polarized optical continuum source is expected, but scaling the Crab Nebula flux up by a factor of 30 leads to a flux level of only $3 \times 10^{-18} \text{ erg cm}^{-2} \text{ s}^{-1} \text{ \AA}^{-1}$ at 6,000 $\text{\AA}$ for a source in

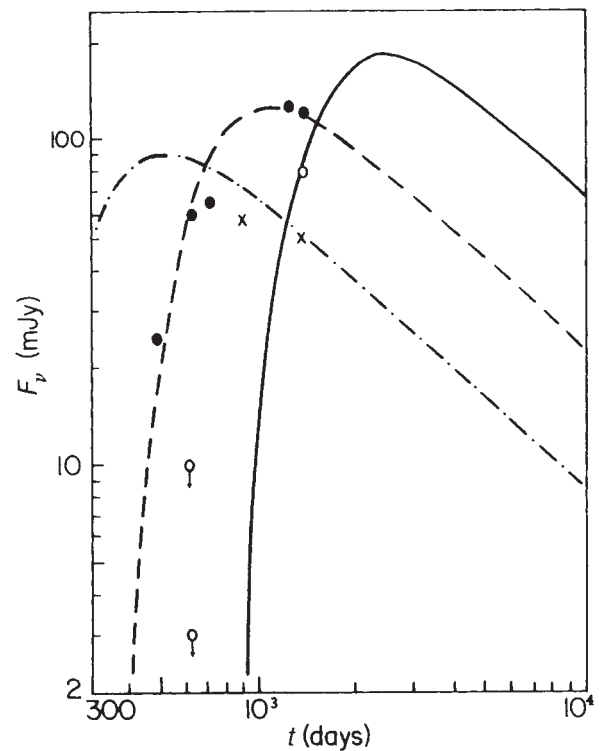


Fig. 1 Comparison of the circumstellar interaction model with the radio flux measurements of SN1986j. The model curves are for wavelengths of 20 cm (solid line), 6 cm (dashed), and 2 cm (dash-dot). The model is from ref. 3 with an electron spectral index of 2.8 and the data are from the compilation in ref. 1 ($\times$, 2 cm; $\bullet$, 6 cm; $\circ$, 20 cm). An explosion in January 1983 is assumed.

NGC891. A young pulsar may have bright pulsed optical emission¹⁵, but the expected luminosity is somewhat below the current detection threshold¹⁶. Finally, we note that the radio and optical emissions from SN 1986j have been modelled as separate physical phenomena, so it is not possible to use a change in radio angular size and an optical velocity to determine the distance to the supernova. The radio and optical properties are more closely related in the model of ref. 5, where the line emission is from a cooling reverse shock wave. If this is the case, the supernova is considerably older than the estimate given here. VLBI observations of the expansion of the radio source should be capable of distinguishing between the models.

I am grateful to N. Bartel and A. Hamilton for discussions. This research was supported in part by the NSF and NASA.

Received 23 June; accepted 5 August 1987.

1. Rupen, M. P., van Gorkom, J. H., Knapp, G. R., Gunn, J. E. & Schneider, D. P. *Astr. J.* (submitted).
2. Chevalier, R. A. *Astrophys. J.* **259**, 302–310 (1982).
3. Chevalier, R. A. *Ann. N.Y. Acad. Sci.* **422**, 215–232 (1984).
4. Weiler, K. W., Sramek, R. A., Panagia, N., van der Hulst, J. M. & Salvati, M. *Astrophys. J.* **301**, 790–812 (1986).
5. Hamilton, A. J. S. *et al.* in *IAU Colloq.* **101** (in the press).
6. Bartel, N., Rupen, M. & Shapiro, I. *IAU Circ. No.* 4292 (1987).
7. Aaronson, M. *et al. Astrophys. J. Suppl. Ser.* **50**, 241–262 (1982).
8. Tammann, G. A. in *Proc. IAU Symp.* **124** (eds Burbidge, G. & Fang, L.-Z.) (Reidel, Dordrecht, in the press).
9. Marscher, A. P. in *Supernovae as Distance Indicators* (ed. Bartel, N.) 130–137 (Springer, Berlin, 1985).
10. Dwek, E. *Astrophys. J.* **274**, 175–183 (1983).
11. Chevalier, R. A. *Astrophys. J.* **258**, 790–797 (1982).
12. Chevalier, R. A. in *Supernovae* (ed. Schramm, D. N.) 53–61 (Reidel, Dordrecht, 1977).
13. Nomoto, K. in *The Crab Nebula and Related Supernova Remnants* (eds Kafatos, M. J. & Henry, R. B. C.) 97–113 (Cambridge University Press, 1985).
14. Uomoto, A. & Kirshner, R. P. *Astrophys. J.* **308**, 685–690 (1986).
15. Pacini, F. *Astrophys. J.* **163**, L17–L19 (1971).
16. Middleton, J. & Kristian, J. *Astrophys. J.* **279**, 157–161 (1984).

Evidence for strong magnetic fields in the inner envelopes of late-type stars

Richard Barvainis*‡, Gordon McIntosh†
& C. Read Predmore†

* National Radio Astronomy Observatory, Edgemont Road,
Charlottesville, Virginia 22901, USA

† Department of Physics and Astronomy,
University of Massachusetts, Amherst,
Massachusetts 01003, USA

Strong SiO masers occur in the inner regions of the stellar envelopes (within a few stellar radii) of variable late-type stars (Miras, semi-regular variables and supergiants), and in the vicinity of IRC2 in Orion, a site of recent star formation. Although linear polarization is a common feature of these masers¹⁻³, previous attempts to detect circular polarization have been inconclusive. Johnson and Clark⁴ claimed detection of 10–20% polarization in Stokes *V* (left-hand circular minus right-hand circular) spectra of the Orion masers at 86.2 GHz. Troland *et al.*¹ failed to confirm this result, while reporting small signals in Stokes *V* spectra for several sources which they attributed to instrumental effects. Here we report sensitive measurements of circular polarization in SiO masers associated with late-type stars, which reveal significantly polarized features at the several per cent level in five of six stars observed. Under the Zeeman splitting hypothesis, these measurements imply magnetic fields in the masing gas in the range 10–100 G. The masers associated with IRC2 in the Orion-KL region were not detected in circular polarization.

The primary motivation for our new search for circular polarization in SiO masers was based on predictions that magnetic fields in the inner envelopes of late-type stars could be as high as 100 G^{5,6}. These predictions resulted from measurements of circular polarization of OH masers in the outer envelopes of supergiant stars, which indicated fields of several mG at the locations of the OH masers (~100 stellar radii). Strong fields near the stars were inferred by assuming conservation of magnetic flux and extrapolating inward. This prompted us to calculate the Zeeman splitting effect for SiO (which, unlike OH, is non-paramagnetic), showing that a field of 10 G should produce measurable circular polarization, given the very sensitive telescope and receiver systems currently available.

The experiment described here eliminates the main source of instrumental contamination in circular polarization measurements, which is the leakage of linear polarization into the *V* spectrum (see below). In contrast to the previous measurements, we observed the $v = 1, J = 1-0$ (where v and J are the molecular vibrational and rotational quantum numbers, respectively) transition at 43.1 GHz rather than the $v = 1, J = 2-1$ transition at 86.2-GHz. The lower-frequency transition is advantageous because Zeeman splitting produces a constant frequency shift between the left- and right-hand circular components of the radiation, independent of the line frequency. This is apparent as a larger velocity shift relative to the line width at lower frequencies, and hence a larger polarization signal in the *V* spectra.

The observations were conducted on 2–5 November 1986, using the 13.7-m antenna of the Five College Radio Astronomy Observatory (FCRAO) under generally clear skies. The single-sideband system temperature (corrected to outside the atmosphere) was typically 500 K using a cooled Schottky mixer receiver. The FCRAO polarimeter consists of a computer-controlled, rotatable quarter-wave plate made of grooved rexolite. During a given five-minute scan, the plate was switched at a rate of 0.5 Hz so that the fast axis was alternately oriented at +45° and –45° to the horizontally aligned linearly polarized

feed, producing left- and right-hand circular polarization sensitivity. The Stokes *V* spectrum, in the absence of instrumental effects, is just the difference between the accumulated spectra in the two orientations. In initial observations at both the FCRAO and the Haystack Observatory, it was found that the *V* spectra measured in this way contained spurious signals due to leakage of linear polarization. This instrumental effect was corrected by modifying the polarimeter to include a rotatable half-wave plate in the beam path. By switching the half-wave plate by 45° (rotating the linear polarization by 90°), the sign of the spurious signal reverses and can be cancelled completely. Another less serious instrumental effect was due to a slight difference between the coupling of the antenna to right and left circular polarizations. The effect on the *V* spectrum is to produce a scaled-down residual of the total flux (Stokes *I*) line profile. The amplitude of the residual was measured to be ~1.0%, and the *V* spectra were corrected accordingly in post-processing. The Stokes parameters for linear polarization, *Q* and *U*, were also measured during the run, and are presented elsewhere⁷.

The masers in Orion were not detected in circular polarization, with 3 σ upper limits to the fractional circular polarization m_c ($= V/I$) of 0.5% at the *I* emission peak. In contrast, the late-type stars do show significant circular polarization. Maser spectra in *I* and *V* for R Cas and VY CMa, typical of the late-type stars observed here, are presented in Fig. 1. The *V* spectra are often complex, but individual circularly polarized components can frequently be identified with total intensity components at the same velocity. Precise fractional circular polarizations of given components are difficult to determine because of component blending, but fractional polarizations accurate to a factor of two or better are readily obtained. The maximum fractional polarizations observed in each source are listed in Table 1. Both right

Table 1 Maximum polarization and magnetic fields of SiO masers

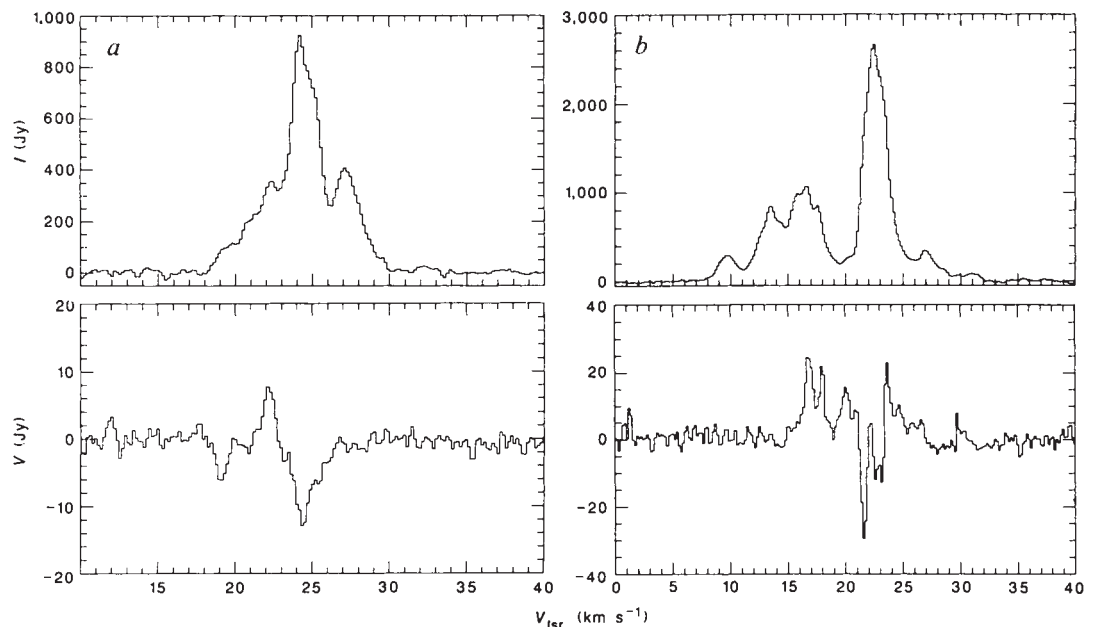
Source	m_c (per cent)	B (G)
NML Tau	<2	<20
Orion	<0.5	<5
R Leo	2.4	24
VY CMa	6.5	65
W Hya	5.0	50
VX Sgr	8.7	87
R Cas	1.5	15

and left circularly polarized features are generally present in the spectrum of a given source, but no unambiguous Zeeman pairs exist in the observations. In this respect, the SiO maser polarization is similar to that of OH masers, which typically show single polarized components unaccompanied by Zeeman counterparts of the opposite sense⁸.

A mechanism first suggested by Cook⁹ has often been invoked to explain the absence of Zeeman pairs in circularly polarized OH masers^{5,6,8}. This mechanism involves accidental matching of magnetic field and velocity gradients within the maser, such that the optical depth and therefore the amplification is greater in one polarization sense than in the other. Detailed radiative transfer calculations have verified that circular polarization in maser line radiation is produced in this model¹⁰, although the magnetic field strength is not directly derivable because the circular polarization depends on the gradient in B rather than on its magnitude. A new mechanism for generating unpaired Zeeman components in masers, requiring a magnetic field and a gradient in the velocity field in the maser (but not requiring a magnetic field gradient), has recently been proposed by Deguchi and Watson¹¹. This model provides a means for estimating the magnetic field strength as a function of m_c , although with uncertainties due both to the dependence of m_c on the angle θ between the velocity gradient and magnetic field directions, and to the fact that the derived field strength is triple-valued for a given absolute magnitude of m_c (see Fig. 2 of ref.

‡ Present address: Haystack Observatory, Westford, Massachusetts 01886, USA.

Fig. 1 Stokes parameters I and V for SiO maser emission from R Cas (a) and VY CMa (b) in the $v=1$, $J=1-0$ transition at 43.1 GHz. The V spectra have been corrected for a spurious 1% residual of the I spectra (see text). Although the up-down pattern seen in V between 21 and 26 km s⁻¹ for R Cas may seem reminiscent of a classical Zeeman S-curve, close comparison of I and V indicates that the up and down features in V are probably associated with two separate and distinct maser components.



11). Here we use this theory and adopt conservative estimates for the field strengths by assuming that the masers are saturated, that $\theta \approx 45^\circ$, and that $\Delta\nu_z < 0.1$ ($\Delta\nu_z$ is the dimensionless ratio of the Zeeman splitting to the local line width). This results in the approximate relation $B \approx 10 \Delta\nu m_c$, where m_c is expressed in per cent, B in gauss, and $\Delta\nu$ is the full width at half-maximum of an individual polarized maser component in km s⁻¹. Using this equation with a typical component line width of 1 km s⁻¹ and the observed maximum fractional polarizations, fields in the approximate range 10–100 G are inferred for the late-type stars (Table 1). Other methods^{5,7} for estimating the magnetic field strength from the circular polarization give similar results. As discussed in ref. 10, the Cook mechanism also requires that $\Delta\nu_z$ be non-negligible, implying order-of-magnitude estimates of B similar to those derived above.

As the masers in our sample show linear as well as circular polarization, limits can be set on free-electron densities in the masing gas from the absence of Faraday depolarization. For a typical magnetic field strength of 50 G, a path length of 10^{14} cm (see below), and assuming that the Faraday rotation is < 1 rad within the maser¹², the derived electron density is $n_e < 15$ cm⁻³.

Cohen *et al.*¹³ have measured narrow circularly polarized components in the 1,612-MHz OH emission from supergiant OH-IR stars, two of which (VY CMa and VX Sgr) are in our source list. They inferred magnetic fields at the locations of the OH masers of $B_{OH} \approx 2$ mG. For VX Sgr, VLBI (very-long-baseline interferometry) maps of the SiO masers¹⁴ show that the maser spots are clustered in a region of radius 30 mas, or 7×10^{14} cm (for a distance of 1.7 kpc). The OH masers lie in a shell at radius 2.1×10^{16} cm (ref. 11). Combining these estimates of the distances of the two maser species with the inferred magnetic field strengths, we find that $B_r \propto r^{-n}$, where n is between 2 and 3, assuming a power-law dependence. Uncertainties in the exponent arise because the calculations of the magnetic field strengths are only approximate, and in any case might not be truly representative of the overall fields in the envelope because the masers might exist in dense clumps (see below). Given these uncertainties, the magnetic field is consistent with a solar-wind type ($\propto r^{-2}$) or a dipole type ($\propto r^{-3}$) configuration.

Knowledge of the physical parameters of SiO masers and their environments is uncertain at present. From VLBI observations and maser models, Moran *et al.*¹⁴ have estimated the characteristic densities and sizes of SiO masers to be $n_H \approx 10^9$ cm⁻³ and $l \approx 10^{14}$ cm respectively; Alcock and Ross¹⁵, using

the same data as well as data presented by Lane¹⁶, have suggested that $n_H \approx 10^{12}$ cm⁻³ and $l \approx 10^{13}$ cm. These authors estimate mean densities in the circumstellar envelope at the radius of the SiO masers of $\sim 10^8$ – 10^9 cm⁻³. Recent work on the outer atmospheres of late-type stars indicates that these stars may have chromospheres out to at least several stellar radii, where the ambient gas has a density of 10^6 – 10^8 cm⁻³ and a temperature of $\sim 8,000$ K (ref. 17). Such a medium may be subject to condensation instabilities, leading to a two-phase medium with the hot gas coexisting with much denser ($n_H \approx 10^9$ cm⁻³) and cooler ($T \approx 10^3$ K) clouds containing molecules and dust; SiO masers may then form in these dense clouds^{18,19}.

The strong magnetic fields we have measured will dominate the energy density and pressure of the gas thermal motions in the maser by a factor $B^2/(8\pi n_H kT) \approx 10^6$ (here and below we assume values of $B = 50$ G, $l = 10^{14}$ cm, $n_H = 10^9$ cm⁻³, $T = 10^3$ K, and an ion number density $n_i = n_e = 10$ in the maser clouds; k is the Boltzmann constant). The adiabatic diffusion timescale of the fields is given by²⁰ $t_{ad} \approx 4\pi\alpha n_H m_H^2 l^2 B^{-2} \approx 10^3$ s, where $\alpha = 5.7 \times 10^{14}$ cm³ s⁻¹ g⁻¹ and m_H is the mass of the hydrogen atom. Since the lifetime of a maser feature is $\sim 10^7$ s, much longer than the diffusion timescale, the fields are not frozen into the masing gas. For $l = 10^{13}$ cm and $n_H = 10^{12}$ cm⁻³, the values advocated by Alcock and Ross¹⁵, the ratio of magnetic to thermal energies is ~ 100 and $t_{ad} \approx 10^4$ s.

Our observations of strong magnetic fields at the sites of SiO masers do not provide direct information on field strengths in the ambient circumstellar gas, if indeed this gas is distinct from that which produces the masers. However, fields considerably weaker than those inferred for the masers will be important in the overall dynamics of the circumstellar envelope. Taking the mechanical energy per unit volume of the maser gas as indicative of that in the near-stellar envelope as a whole, we can estimate the minimum field strengths required for magnetic energy to dominate the energy in gas motions. The velocities of maser features are typically spread over 10 km s⁻¹, and for maser densities of $n_H = 10^9$ cm⁻³, $\epsilon_{mech} = \frac{1}{2} m_H n_H v^2 \approx 10^{-3}$ erg cm⁻³. The magnetic energy density will be larger than this if $B \geq 0.1$ G. Overall field strengths of this magnitude certainly seem plausible, if for no other reason than that the strong fields in the masers are not frozen in, and some amount of diffusion of the field lines into the ambient gas probably takes place. It would appear that magnetic fields must be considered in theories for the expansion dynamics of circumstellar envelopes in late-type

stars. Indeed, mechanisms for lifting material from the star using energy deposited by Alfvén waves²¹ may take on added relevance.

The Five College Radio Astronomy Observatory is operated by Associated Universities Inc., with support from NSF and with permission of the Metropolitan District Commission, Commonwealth of Massachusetts. Radio Astronomy at the Haystack Observatory of the Northeast Radio Observatory Corporation is supported by NSF.

Received 9 June; accepted 12 August 1987.

1. Troland, T. H., Heiles, C., Johnson, D. R. & Clark, F. O. *Astrophys. J.* **232**, 143–157 (1979).
2. Clark, F. O., Troland, T. H. & Johnson, D. R. *Astrophys. J.* **261**, 569–575 (1982).
3. Barvainis, R. E. & Predmore, C. R. *Astrophys. J.* **288**, 694–702 (1985).
4. Johnson, D. R. & Clark, F. O. *Astrophys. J.* **197**, L69–L72 (1975).
5. Reid, M. J. *et al. Astrophys. J.* **227**, L89–L91 (1979).
6. Claussen, M. J. & Fix, J. D. *Astrophys. J.* **263**, 153–159 (1982).
7. McIntosh, G. thesis, Univ. of Massachusetts (1987).
8. Reid, M. J. & Moran, J. M. *A. Rev. Astr. Astrophys.* **19**, 231–276 (1981).
9. Cook, A. H. *Celestial Masers* (Cambridge University Press, 1977).
10. Deguchi, S. & Watson, W. D. in *Star Formation in Galaxies* (ed. Lonsdale Persson, C. J.) 87–90 (NASA, Washington DC, 1987).
11. Deguchi, S. & Watson, W. D. *Astrophys. J.* **300**, L15–L17 (1986).
12. Goldreich, P., Keeley, D. A. & Kwan, J. Y. *Astrophys. J.* **179**, 111–134 (1973).
13. Cohen, R. J. *et al. Mon. Not. R. astr. Soc.* **225**, 491–498 (1987).
14. Moran, J. M. *et al. Astrophys. J.* **231**, L67–L69 (1979).
15. Alcock, C. & Ross, R. R. *Astrophys. J.* **310**, 838–841 (1986).
16. Lane, A. P. thesis, Univ. Massachusetts, Amherst (1982).
17. Linsky, J. L. *Irish astr. J.* **17**, 343–351 (1986).
18. Stencel, R. E., Carpenter, K. G. & Hagen, W. *Astrophys. J.* **308**, 859–867 (1986).
19. Muchmore, D. O., Nuth, J. A. & Stencel, R. E. *Astrophys. J.* **315**, L141–L146 (1986).
20. Mestel, L. in *Protostars and Planets II* (eds Black, D. C. & Matthews, M. S.) 320–339 (University of Arizona Press, Tucson, 1986).
21. Hartmann, L. & MacGregor, K. B. *Astrophys. J.* **242**, 260–282 (1980).

Discovery of organic grains in comet Wilson

D. A. Allen* & D. T. Wickramasinghe†

* Anglo-Australian Observatory, Epping, NSW 2121, Australia

† Department of Mathematics, Australian National University, Canberra, ACT 2600, Australia

Organic materials have been shown to be a major component of the solid grains in long path lengths through the Galaxy^{1–3}. Their identification was through the C–H bond stretching vibration transition near 3.4 μm , which in grains is seen as a broad, weakly structured absorption feature on the radiation of bright background sources. A similar spectral feature was seen in emission in comet Halley^{4–7}. Direct measurement of the grain composition near the nucleus by mass spectrometry^{8,9} confirmed the organic composition of much of the material shed by comet Halley. There, however, the surface appears to have been heavily processed by long exposure to sunlight, and it remains unclear whether the organic material represented a release of pristine grains from within the nucleus, or the stripping from the surface of the processed material. Here we report the detection of a similar, but distinct, emission feature in comet Wilson, a comet probably making its first visit to the solar vicinity.

The observations were made with the 3.9-m Anglo-Australian Telescope (AAT), using the infrared photometer-spectrometer in its spectroscopic mode. The data acquisition was similar to that used for comet Halley⁵ with a single exception. Because comet Wilson was considerably fainter than Halley, we used the focal plane chopper available in the instrument to difference rapidly between the comet and a nearby portion of sky. When observing Halley, we were able to use a more distant sky reference, sampled more rarely, but we were then sensitive to rapid drifts in the sky signal.

The separation of the signal and reference beams was 12 arc s, and by additionally moving the telescope through the same

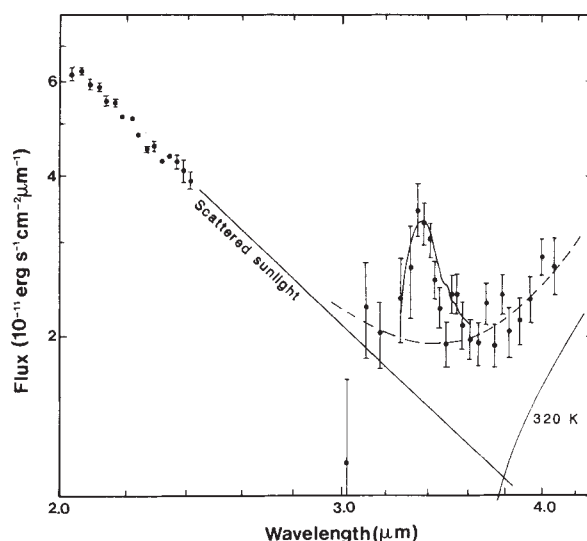


Fig. 1 Infrared spectrum of comet Wilson. The expected continua (scattered sunlight and thermal radiation) add to make the broken line. The emission feature seen in comet Halley, added to this continuum, is shown as a thin solid line.

distance we sampled the difference between the signal at the nucleus and at positions symmetrically placed 12 arc s to the north and south. We therefore subtracted a small portion of the coma from the nucleus. In comet Halley, the spectrum in corresponding portions of the coma was similar to that from the nucleus⁷, so we think it unlikely that our data were influenced by any more than a uniform weakening, estimated at ~20%.

Data were taken on the night of 11 May UT. We used a 5-arc s circular aperture, and centred our entrance aperture on the optically brightest position. As for comet Halley, the raw spectral resolution, $\lambda/\Delta\lambda$, was ~90, but we have binned the data to improve the signal-to-noise ratio, and the data presented below have an effective resolution of 60.

A nearby bright star, BS2934, was used as a secondary standard for the principal wavelength range observed, 3.0–4.1 μm , and was observed on four occasions on either side and during periods of observation of the comet. In this way, we averaged any variations in the atmospheric transmission through the highly attenuated portion to 3.35 μm . BS2934 was calibrated by observations of BS2882, one of the AAT spectrophotometric standards¹⁰, and is adequately represented by the same spectral function 31 times brighter. Observations of the comet were also made over the range 2.0–2.4 μm to constrain the continuum shape.

Figure 1 shows the data in much the same format that we presented our results on comet Halley⁵. As for comet Halley, we have fitted a continuum by adding together one component, falling to longer wavelength and attributed to scattered sunlight, and a second, rising component due to thermal emission from the dust grains. The continuum was determined by data points outside the waveband of the emission feature seen in comet Halley, with the exception of the clearly discrepant point at 3.03 μm .

The continuum shows some differences from that of comet Halley. Comet Wilson lay at a slightly greater distance from the Sun, enough to have reduced the colour temperature of the dust grains from 350 K, found in Halley, to 340 K. We seem to require a lower temperature, and have used 320 K; this suggests that the grains are somewhat larger than in Halley. The opposite can be inferred from the bluer colour of the scattered sunlight. But the data are barely adequate to derive profound conclusions, and in particular a temperature of 340 K cannot be precluded.

An emission band is clearly seen near 3.4 μm . We have attempted to fit the band by the feature we observed in comet

Halley. The use of our Halley data was purely for convenience: the data of ref. 7 are essentially identical. We subtracted the fitted continuum from the Halley data, scaled the resulting feature appropriately, and added it to the derived Wilson continuum.

The envelope of the emission band in comet Wilson closely matches that of comet Halley, and its intensity relative to the local continuum lies within the range of variability seen in comet Halley. It is known from the Giotto and Vega data^{8,9} that the material producing the 3.4- μm feature in comet Halley was in the solid phase. We cannot be so sure in this case. The narrowness of the two peaks might be seen as evidence that we have measured material in the gas phase. Arguments based on the predicted intensity of gaseous emission lines in comet Halley⁶ can be transferred to comet Wilson, and give us confidence that the emission arises in small grains.

The black surface of comet Halley's nucleus^{11,12} is interpreted as internal dust grains processed by exposure to sunlight. The violent outgassing observed from comet Halley must strip off some of its surface material. The darkness of the grains supports this view, especially as there is a suggestion that the dust albedo rises during major jet activity¹³. Thus, its infrared spectra may be partly due to processed material.

Comet Wilson has a hyperbolic orbit (eccentricity 1.000323, see ref. 14), suggesting that this is its first visit to the solar neighbourhood. Therefore, we should be viewing the most pristine version yet encountered of the organic grains which condense in the sheltered environments of star-forming regions. Any differences between comets Wilson and Halley might reflect local conditions in their respective birthplaces. On the other hand, they may also arise because of their diverse histories.

We suspect one important difference. In comet Wilson the emission band appears to break into two components, centred near 3.38 and 3.52 μm . At a resolution of 90 in comet Halley the band showed only an inflection between these two wavelengths^{5,7}; at higher spectral resolution Baas *et al.*⁶ found two distinct peaks, at 3.36 and 3.52 μm . It is difficult to interpret the data of Baas *et al.* because no continuum fit was performed, but they claim that the emission feature fell completely to the continuum level between the two peaks. Our data on comet Wilson suggest that the same two emission peaks were present. As our present observations had lower spectral resolution than any of the data taken on comet Halley, each peak would seem to be narrower than those in comet Halley.

In the interstellar environment, chemical modification by starlight will run its full course. It is well established that infrared bands in relatively simple organic solids are modified by exposure to ultraviolet radiation, usually in the sense that the spectral features become more diffuse¹⁵. The same progression is suggested by our data: comet Wilson shows the sharpest features; comet Halley represents an intermediate stage, and the ambient material on sight lines through the Galaxy^{16,17} has the most diffuse band. We have already noted⁵ that the best match to the feature in comet Halley is seen in the ρ Ophiuchi dark cloud, an environment which probably offers shielding intermediate between that of a pristine comet and interstellar gas clouds.

Therefore we consider it vital to obtain the best possible data on comets like Wilson which are believed to be making their first pass of the Sun. It is regrettable that comet Wilson proved fainter than originally anticipated, and that cloud further curtailed our observations. The data presented here are not as high quality as we had hoped; fortunately, observations of comet Wilson made nearly two weeks later have been reported by Brooke *et al.*¹⁸, who noted 2.8- and 3.4- μm bands which "closely resemble the unidentified emission features discovered in" comet Halley.

The band shape we observe does not appear to match the feature in the few biological materials so far studied¹⁹, and therefore places some additional constraints on the Hoyle-

Wickramasinghe biological model of cometary dust²⁰. We recommend those exploring laboratory identifications of the organic grains to seek materials which match the profile of the 3.4- μm band in comet Wilson and which, after exposure to the conditions expected in the interstellar radiation field, develop band structures resembling those towards the Galactic centre.

Received 12 June; accepted 25 August 1987.

1. Wickramasinghe, D. T. & Allen, D. A. *Nature* **287**, 518-519 (1980).
2. Allen, D. A. & Wickramasinghe, D. T. *Nature* **294**, 239-240 (1981).
3. Wickramasinghe, D. T. & Allen, D. A. *Astrophys. Space Sci.* **97**, 369-378 (1983).
4. Combes, M. *et al.* *Nature* **321**, 266-268 (1986).
5. Wickramasinghe, D. T. & Allen, D. A. *Nature* **323**, 44-46 (1986).
6. Baas, F., Geballe, T. R. & Walther, D. M. *Astrophys. J.* **311**, L97-L101 (1986).
7. Knacke, R. F., Brooke, T. Y. & Joyce, R. R. *Astrophys. J.* **310**, L49-L53 (1986).
8. Kissel, J. *et al.* *Nature* **321**, 280-282 (1986).
9. Kissel, J. *et al.* *Nature* **321**, 336-337 (1986).
10. Allen, D. A. & Cragg, T. A. *Mon. Not. R. astr. Soc.* **203**, 777-783 (1983).
11. Sagdeev, R. Z. *et al.* *Nature* **321**, 262-266 (1986).
12. Keller, H. U. *et al.* *Nature* **321**, 320-326 (1986).
13. Tokunaga, A. T., Golisch, W. F., Griep, D. M., Kaminski, C. D. & Hanner, M. S. *Astr. J.* **92**, 1183-1190 (1986).
14. Green, D. W. E. *Int. astr. Union Circ.* No. 4364 (1987).
15. d'Hendecourt, L. B., Allamandola, L. J., Grim, R. J. A. & Greenberg, J. M. *Astr. Astrophys.* **158**, 119-134 (1986).
16. Jones, T. J., Hyland, A. R. & Allen, D. A. *Mon. Not. R. astr. Soc.* **205**, 187-190 (1983).
17. Butchart, I., McFadzean, A. D., Whittet, D. C. B., Geballe, T. R. & Greenberg, J. M. *Astr. Astrophys.* **154**, L5-L7 (1986).
18. Brooke, T. Y., Knacke, R. F., Owen, T. & Tokunaga, A. T. *IAU Circ.* No. 4399 (1987).
19. Hoyle, F., Wickramasinghe, N. C., Al-Mufti, S., Olavesen, A. H. & Wickramasinghe, D. T. *Astrophys. Space Sci.* **83**, 405-409 (1982).
20. Hoyle, F. & Wickramasinghe, N. C. *Living comets* (University College Cardiff Press, 1985).

An important uncertainty in coupled chlorine-carbon dioxide studies of atmospheric ozone modification

R. S. Eckman*, J. D. Haigh† & J. A. Pyle*

* Department of Physical Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EP, UK

† Centre for Remote Sensing, The Blackett Laboratory, Imperial College of Science and Technology, London SW7 2BZ, UK

There has been considerable interest recently in changes in the composition of the stratosphere following the suggestion by Molina and Rowland¹ that certain fluorocarbons could ultimately lead to ozone destruction. This interest has increased since the discovery in 1985 of large ozone decreases over Antarctica². It is now known that many species have the potential to modify the stratospheric ozone distribution either directly, by photochemical processes, or indirectly, through changes in atmospheric temperature by modification of the transmission of infrared radiation. Here we present calculations using a two-dimensional atmospheric model of the effect on stratospheric ozone of increases in the atmospheric abundances of CO₂ and various chlorine-containing compounds. We have carried out such calculations previously³ but changes in the kinetic data used in the model have led to very significant differences in the calculated ozone modification. In particular, the lower stratosphere, a difficult region to model, now plays an even more crucial role in the vertically integrated ozone depletion. We believe that uncertainties in the thermal response of the lower stratosphere represent a significant limitation in our ability to predict future states of the middle atmosphere.

The two-dimensional atmospheric model extends from the ground to ~95 km and from pole to pole. The resolution is ~9.5° in the horizontal and 0.5 pressure scale heights (~3.5 km) in the vertical. An important feature of this model is that the circulations are calculated from the driving forces of radiative heating and eddy heat and momentum forcing (with the latter specified from satellite observations). The model can include the important feedback between ozone amount, temperature and radiative heating and cooling. The modelled ozone values

are used in the calculation of the solar heating (but not in the long-wave cooling by the 9.6- μm band which uses a standard ozone profile and the cooling-to-space approximation). The most important cooling term is that due to the 15 μm band of CO_2 and in our treatment the heating profile H , at each model level is related to the temperature or Planck function profile, B , by $H = C B$. The matrix C^{4-6} describes the atmospheric transmission between different levels, which for a uniformly mixed gas such as CO_2 needs to be calculated just once, assuming that the transmission is not strongly dependent on the temperature. Deviations from local thermodynamic equilibrium are included in the calculation of C .

As in our previous calculations which considered increasing CO_2 , the effect on tropospheric temperatures are included by changing the sea surface temperatures. For a doubling of CO_2 we have assumed a uniform 2 K temperature increase, consistent with cases discussed in NASA/WMO⁷. The lower stratosphere is again assumed to be in radiative equilibrium. While this is clearly unsatisfactory we again offer the *a posteriori* justification that our calculated temperature changes agree qualitatively with three-dimensional simulations^{8,9} of the troposphere and stratosphere. Furthermore, we have also performed calculations using a newtonian cooling scheme in the lower stratosphere. These calculations produce essentially the same results as reported here.

The model has been described previously^{3,10}. It has been used to study the effect of chlorine compounds on stratospheric composition¹¹. The first two-dimensional model studies of the effect of the increasing levels of CO_2 on ozone¹² used this model. Simultaneous chlorine and CO_2 calculations have also been described³. Changes to the model compared to these previous studies are an extension of the chemical schemes to include N_2O_5 and HOCl (and a number of additional chlorine source gases). Here continuity equations are solved for the following species or groups of species: $\text{O}(^1\text{D})$, O and O_3 ; NO , N , ClONO_2 , NO_2 , N_2O_5 ; HNO_3 ; Cl , ClO , HOCl , HCl and ClONO_2 ; HNO_4 ; OH , H , HO_2 (with transport ignored) and H_2O_2 . The kinetic data uses the updated values recommended in various recent reports^{7,13}. The matrices C also include some small improvements on previous studies.

We have carried out five calculations. The control run, A, has a CO_2 mixing ratio of 330 p.p.b.v. and just over 2 p.p.b.v. of odd chlorine ($\text{Cl} + \text{HCl} + \text{ClONO}_2 + \text{ClO} + \text{HOCl}$) in the upper stratosphere, roughly representative of the present day. Four steady-state calculations have also been performed. First, run B is similar to run A but with the odd chlorine levels increased by a factor of 3. Second, in run C, the abundance of CO_2 is doubled (to 660 p.p.m.v.). Third, run D, includes both increased chlorine and CO_2 as in the two previous runs. Finally, another coupled perturbation, run E, was run with the level of chlorine increased by a factor of 4 to ~ 8.5 p.p.b.v. and with CO_2 at

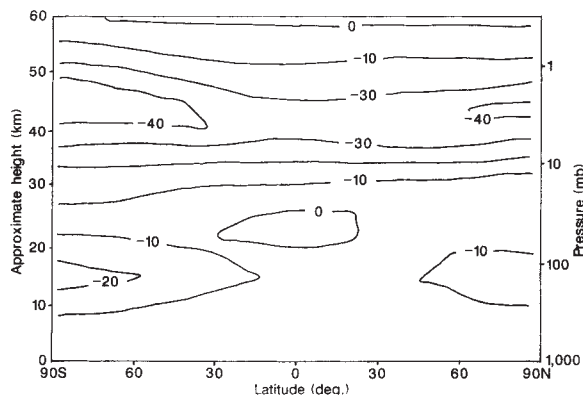


Fig. 1 Percentage difference in ozone mixing ratio in April due to tripling chlorine.

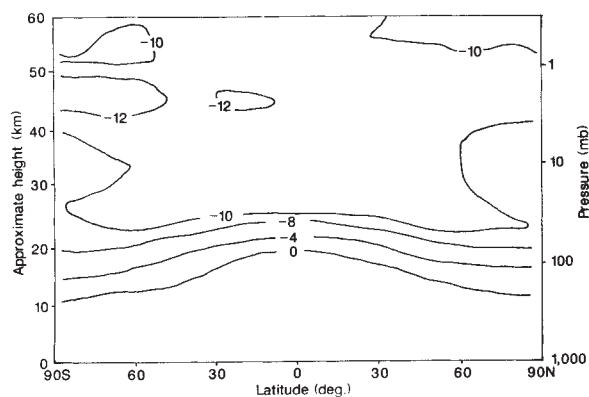


Fig. 2 Cross-section of temperature difference (K) in April due to doubling CO_2 .

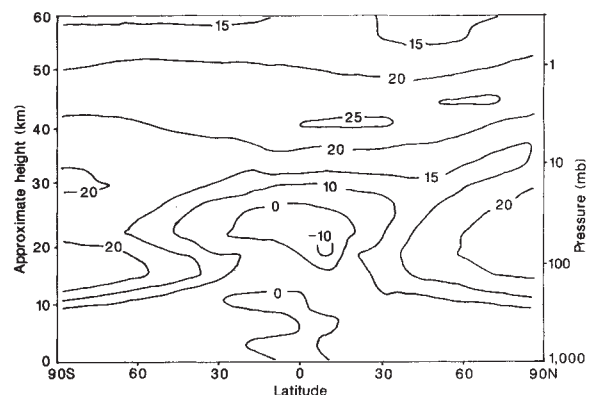


Fig. 3 Percentage change in ozone mixing ratio in April due to doubling CO_2 .

660 p.p.m.v. These represent a limited range of possibilities and exclude changes in other halogen-containing compounds as well as CH_4 and N_2O which are currently increasing. The 8 p.p.b.v. case was considered in the recent NASA/WMO report⁷.

Runs B and C produce results similar to those previously calculated^{3,11,12}. For example, increasing odd chlorine in the stratosphere has led to a significant depletion of ozone at ~ 40 km with smaller changes at lower altitudes (see Fig. 1). There is an area of 'self-healing' in the equatorial lower stratosphere (decreased O_3 in the upper stratosphere leading to increased downward transmission of ultraviolet radiation). The percentage reduction in the vertically integrated ozone amount (column ozone) is smallest in low latitudes and shows significant seasonal variations in high latitudes (not shown).

Figure 2 shows the temperature difference calculated between runs C and A. The result is in good agreement with our previous calculations^{3,12}. This is not surprising because the matrices C have changed only slightly. Figure 3 shows the ozone change calculated in run C with respect to run A. The reduced temperatures have led to an increase in ozone at ~ 40 km and in high latitudes of the lower stratosphere. There is an area of reverse self-healing (ozone reduction) in the equatorial lower stratosphere. In general, the stratospheric results also agree well with our previous results but some differences are found in the troposphere. Previously we had found small ozone decreases here. In these runs there are small ozone increases, with the area of decrease confined to equatorial regions. The change in the column shows largest percentage increases in high latitudes (not shown).

Turning to the coupled perturbation, Fig. 4a shows the percentage ozone change between runs D and A. The upper stratospheric changes are in accord with earlier work³. The ozone

depletion is larger than would be found if the two independent perturbations were added. The nonlinearity arises because the addition of chlorine modifies the temperature dependence of ozone. The ozone changes calculated in the lower stratosphere are radically different from those found before. Ozone has decreased in low latitudes but there is a significant increase, of $\sim 10\%$, in high latitudes found throughout the year. This important feature has not been reported previously. It is extremely unlikely that it has been produced by a self-healing mechanism which would be most evident in the tropics. Neither is it likely that the increased levels of ozone are due to transport as the ozone increases are almost entirely found in high latitudes; there is no pool of increased ozone available for transport from low latitudes. It must be concluded then that the ozone change is produced *in situ*.

As the temperature changes are similar to our earlier work the ozone changes must result from the improvement in our photochemical scheme: either the introduction of new species or the inclusion of more recent kinetic and photochemical data. Figure 4b provides the clue to the likely cause of our new result. It shows the difference between the OH concentrations found in runs B and D. In run D, which differs from run B in having lower temperatures in the stratosphere, we find much lower concentrations of OH, by $\leq 50\%$ in high latitudes at ~ 25 km. These changes result primarily from the use of new kinetic data for the reactions $\text{OH} + \text{HO}_2$ and $\text{OH} + \text{HNO}_3$ which are the major sinks for hydrogen-containing radicals in the lower stratosphere. The rate constants for these reactions have a significant negative activation energy so that at lower temperatures (as in run D) the reactions proceed at a greater rate. Thus, in run D the concentrations of hydrogen radicals are reduced in the high-latitude lower stratosphere. Comparison of the changes in other constituents confirms our hypothesis. There is no change, for example, in the total amount of odd nitrogen and only small changes in the ratio of NO to NO_2 . The increase in ozone in run D appears to result from a reduction in ozone destruction catalysed by the hydrogen radicals.

To confirm these ideas we have looked at the contribution to ozone destruction by the various catalytic cycles in the different runs. The hydrogen chemistry makes the largest contribution to ozone destruction in the low stratosphere at ≤ 20 km, with nitrogen chemistry rapidly increasing in importance above. Between runs B and D the ozone loss due to reactions of OH and HO_2 decreases by $> 30\%$ at ~ 50 mb, indicating that this is indeed the main cause of the ozone increase in run D. The transition between positive and negative ozone changes near 10 mb shown in Fig. 4a is in accord with the decreased odd hydrogen catalysed loss.

Figure 4c shows how these changes in ozone translate into changes in the vertical column. A latitudinal behaviour, completely different from all previous studies, is found with the largest percentage depletions now found in equatorial latitudes with ozone increases predicted for high latitudes. When run E (with 8.5 p.p.b.v. odd chlorine and 660 p.p.m.v. CO_2) is compared with run A, column ozone depletions are found at all latitudes (not shown) and the gradient is similar to that shown in Fig. 4c, with largest depletions found in low latitudes.

These results indicate that model calculations of ozone depletion (or increase) depend sensitively on the temperature perturbations in the lower stratosphere. This is a particularly difficult region for radiation calculations with a small net balance of several effects. Furthermore, the radiative relaxation time constant is long and dynamical adjustments will be very important. The model used here includes the important feedback between ozone concentration, temperature and radiative heating rates and incorporates more sophisticated radiative transfer calculations than used in other two-dimensional photochemical models. Nevertheless, the temperatures calculated here should not be regarded as definitive, but merely as indicating an important uncertainty. For example, using the same model¹⁴ it has

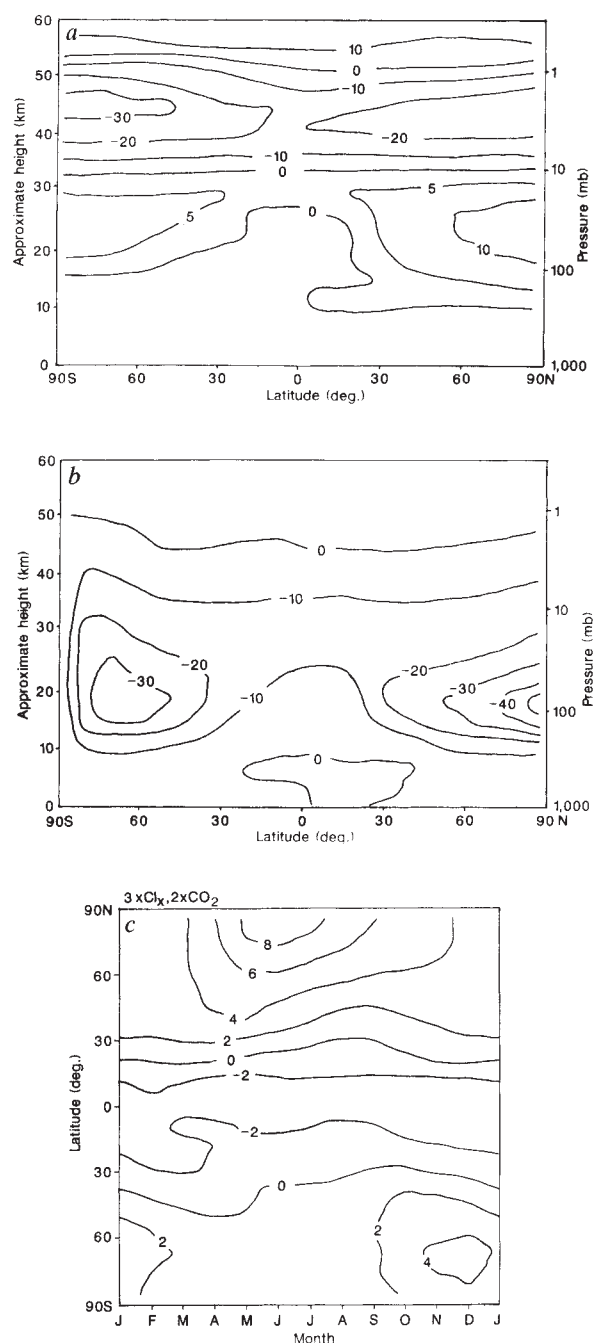


Fig. 4 a, Percentage change in ozone mixing ratio in April for the simultaneous perturbations. b, Percentage difference in OH mixing ratio in April between the simultaneous perturbations and tripling chlorine only. c, Latitude-time section of the percentage difference in total ozone for the simultaneous perturbations.

previously been shown that further modifications to the radiation schemes, including relaxation of the assumption of radiative equilibrium in the lower stratosphere, significantly affects the calculated temperature and hence ozone in both present and perturbed atmospheric simulations. That study predicted much smaller changes in temperature at ~ 25 km for a doubling of CO_2 than reported here. Three-dimensional model results⁸ are intermediate between these two calculations. None of these studies included the radiative effects of other gases (for example, the fluorocarbons, CH_4 and N_2O) that we expect to contribute to a cooling of the lower stratosphere. It seems important, therefore, to make a simulation of the temperature changes as accurately as possible and, in future, we plan to make more

sophisticated radiation calculations at all altitudes and also to include the effects of the other radiatively active gases.

The work at the University of Cambridge was supported by the UK Department of the Environment. We thank Mr P. R. Austin for his efforts in the preliminary stages of this work.

Received 20 May; accepted 7 August 1987.

1. Molina, M. J. & Rowland, F. S. *Nature* **249**, 810-812 (1974).
2. Farman, J. C., Gardiner, B. G. & Shanklin, J. D. *Nature* **315**, 207-210 (1985).
3. Haigh, J. D. & Pyle, J. A. *Q. Jl R. met. Soc.* **108**, 551-574 (1982).
4. Curtis, A. R. *Proc. R. Soc. A* **236**, 156-159 (1956).
5. Williams, A. P. thesis, Univ. Oxford (1971).
6. Houghton, J. T. *The Physics of Atmospheres* 63 (Cambridge University Press, 1977).
7. NASA/WMO *Atmospheric Ozone 1985* (World Met. Org., Geneva, 1986).
8. Fels, S. B., Mahlman, J. D., Schwarzkopf, M. D. & Sinclair, R. W. *J. atmos. Sci.* **37**, 2265-2297 (1980).
9. Manabe, S. & Wetherald, R. T. *J. atmos. Sci.* **37**, 99-118 (1980).
10. Harwood, R. S. & Pyle, J. A. *Q. Jl R. met. Soc.* **101**, 723-748 (1975).
11. Pyle, J. A. *Pure appl. Geophys.* **118**, 355-377 (1980).
12. Haigh, J. D. & Pyle, J. A. *Nature* **279**, 222-224 (1979).
13. JPL/NASA *Chemical Kinetics and Photochemical Data for Use in Stratospheric Modelling, Evaluation No. 7* *Jet Propul. Lab. Publ.* 85-37 (Jet Propulsion Laboratory, Pasadena, California, 1985).
14. Haigh, J. D. *Q. Jl R. met. Soc.* **110**, 167-185 (1984).

Sustained chemical waves in an annular gel reactor: a chemical pinwheel

Zoltan Noszticzius*, W. Horsthemke*,
W. D. McCormick, Harry L. Swinney & W. Y. Tam

Center for Nonlinear Dynamics and Department of Physics,
The University of Texas, Austin, Texas 78712, USA

Past experiments on the formation of chemical spatial structures¹⁻⁴ (mainly chemical waves) have been conducted in closed, batch systems such as Petri dishes⁵⁻¹⁵, and not in open chemostats such as those usually employed in the study of temporal structures. As a consequence, the observed spatial patterns change uncontrollably and ultimately decay as the mixture evolves towards thermodynamic equilibrium. We have overcome this problem in a novel type of open reactor: the reaction occurs in an annular inert gel that is fed at the inner and outer rims. The reaction that we have studied is the Belousov-Zhabotinskii (BZ) reaction, which has become the paradigm for studies of spatial and temporal structures in nonlinear systems. In the ring reactor we observe sustained travelling azimuthal waves for a wide range of conditions.

The ring reactor, shown diagrammatically in Fig. 1, surmounts two other problems encountered in previous experiments: (1) convective motion¹⁶⁻²⁰, generated by concentration gradients or other instability mechanisms, cannot occur in the gel, and (2) the gel also inhibits the formation of the bubbles that plague studies of pattern formation in the BZ reaction²¹.

The polyacrylamide gel in the ring reactor stretches 1-2 mm beyond the inner and outer edges of the Plexiglas rings that hold it in place (see Fig. 1). The sulphuric acid and potassium bromate components of the BZ reaction are in the outer compartment, while malonic acid with the ferroin catalyst is in the inner compartment. The two solutions are separated by the gel.

After the start of an experiment, a red concentric front of ferroin can be seen to be advancing slowly toward the outer edge of the annulus. The front is rather sharp at the beginning (Fig. 2a), but gradually becomes more diffuse (Fig. 2b). After 2-3 hours, the circular symmetry of the front starts to break,

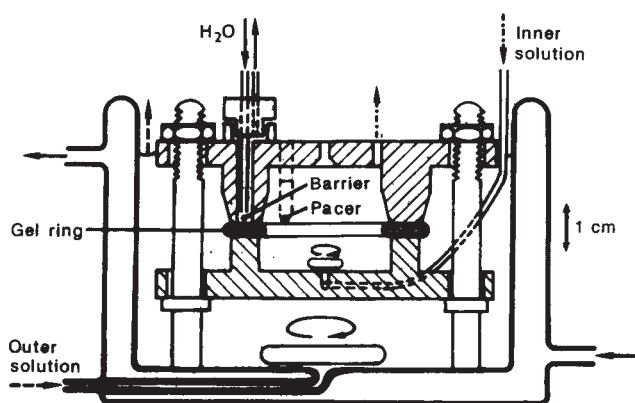


Fig. 1 Cross-sectional view of the apparatus. The 2-mm-thick gel, placed between two Plexiglas rings of inner radius 27 mm and outer radius 38 mm, separates magnetically stirred inner and outer solutions. The whole structure including the gel ring is in a thermostated (29 °C) glass reactor containing the outer solution. Waves travelling in the annulus can be observed through the transparent top. The Plexiglas ring below the gel is painted with a white paint (TiO₂ mixed with Plexiglas solution in CHCl₃) and covered with a thin protective layer of Plexiglas film; the white background produced this way ensures good visibility and contrast. The polyacrylamide gel was prepared from the following aqueous solutions: a, 20 g acrylamide per 100 ml; b, 0.5 g N,N'-methylene-bis-acrylamide per 25 ml; c, 2 g ammonium persulfate per 10 ml and d, 3 g triethanolamine per 10 ml. b was filtered before use, but the other solutions were clear without filtration. The gel was formed by first mixing 10 ml a, 15 drops b, 10 drops c, and 10 drops d together vigorously for 10 s. Then the mixture was poured into a Plexiglas form, and after 10 min of polymerization it was washed in distilled water for 40 min (the gels swelled by about 20% during the washing).

and more and more local irregularities are observed (Fig. 2c). Finally, a state develops with one or more pacemaker centres (wave sources) and with the same number of annihilation points (sinks); Fig. 2d shows the simplest case with one source and one sink. The number and the frequency of the pacemakers vary randomly, in agreement with Winfree's observations²². In general, the numbers of clockwise and counterclockwise propagating waves are different, but after an initial period of typically 8 h, the difference between these numbers becomes constant because the waves are produced and eliminated in clockwise-counterclockwise pairs.

While the behaviour of the spontaneously appearing chemical waves in the system is rather idiosyncratic, we have found after an appropriate perturbation the system can display stable rotating waves. The system is perturbed using a technique proposed by Winfree²³ on a theoretical basis, in analogy with some neurophysiological experiments. An artificial pacemaker is created by filling the bore (the pacemaker in Fig. 1) with a solution of 0.5M H₂SO₄ and 0.5 M KBrO₃. This perturbant diffuses into the gel through a 0.3-mm diameter hole at the conical end of the bore. To stop the pacemaker, the bore is washed with distilled water. To set up rotating waves, we need to eliminate say, for example, the counterclockwise waves. This is done by setting up a wave barrier close to the pacemaker in the counterclockwise direction. The barrier is a region in the gel in which the reagents have been diluted by a slow continuous flow (~2 ml min⁻¹) of distilled water through the bore (the barrier in Fig. 1). When not in use, both bores are closed with glass stoppers which are removed only when photographs are taken.

Figure 3 shows the rotating chemical waves initiated in the way just described. For the sake of clarity, the pacemaker is inserted first; its effect can be seen in Fig. 3a. Then the barrier is activated; its effect is shown in Fig. 3b. The barrier is not

* Present address: Institute of Physics, Technical University of Budapest, 1521 Budapest, Hungary (Z.N.); W.H. is also at the Center for Studies in Statistical Mechanics, The University of Texas, Austin, Texas 78712, USA.

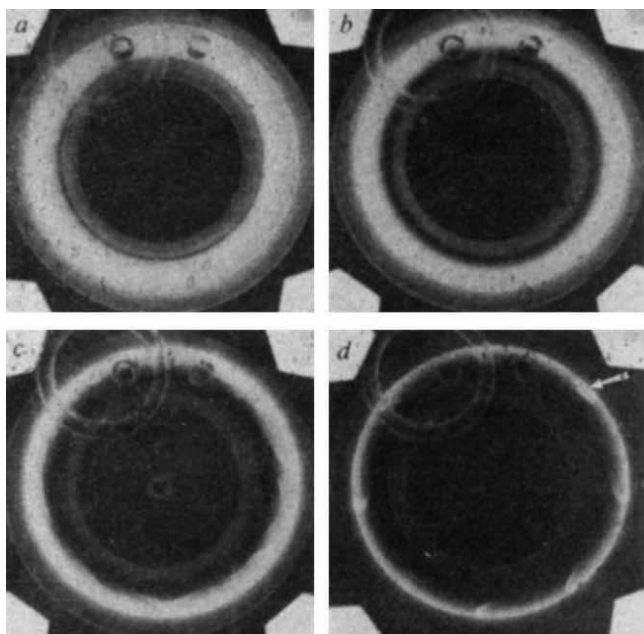


Fig. 2 Temporal evolution of a pattern in the gel reactor. *a*, Initial sharp axisymmetric front at $t = 0.25$ h. *b*, Diffuse axisymmetric front at $t = 1$ h. *c*, Irregular front at $t = 3$ h. *d*, Post-transient state with one source (indicated by the arrow) and one sink (opposite the source) at $t = 12$ h. The concentrations in the outer solution were $[\text{KBrO}_3] = 0.15$ M and $[\text{H}_2\text{SO}_4] = 0.15$ M; in the inner solution, $[\text{CH}_2(\text{COOH})_2] = 0.1$ M and $[\text{ferroin}] = 6 \times 10^{-3}$ M.

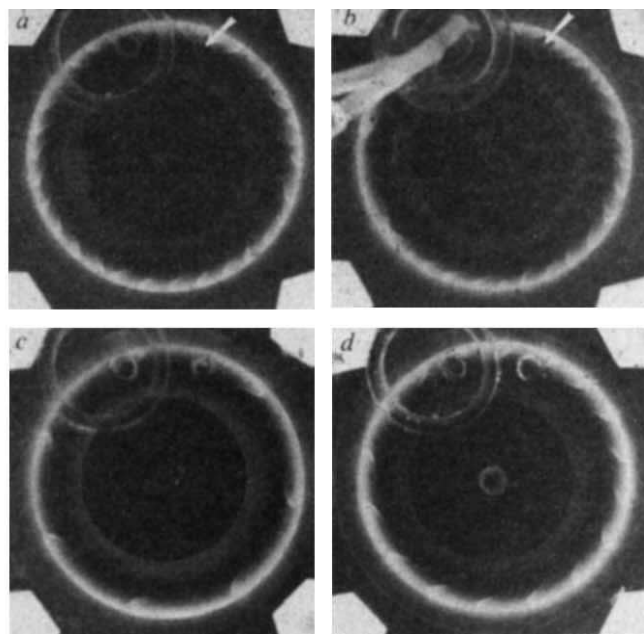


Fig. 3 *a*, Effect of an artificial pacemaker in the absence of the barrier. *b*, Effect of the artificial pacemaker and the wavebarrier. *c*, A rotating chemical structure—a pinwheel—with 7 waves. *d*, A pinwheel with 17 waves. The arrows in *a* and *b* indicate the source.

perfect; every second or third wave goes through. But this still produces an overwhelming excess of clockwise waves after a while and is thus sufficient for our purposes. If we remove the barrier and stop the artificial pacemaker, the production of waves is stopped and the few counterclockwise waves will be annihilated along with an equal number of clockwise waves. Only clockwise rotating waves then remain in the annulus. If the barrier is still active, it will now decrease the number of clockwise waves. In this way the total number of clockwise waves can be varied. After both the pacemaker and the barrier are removed, the initially unevenly spaced waves relax within 3–5 h to a state with equal gaps between the wavefronts.

The uniform wavelength of the post-transient states, illustrated in Fig. 3*c* and *d* for states with 7 and 17 waves respectively, is a consequence of the dependence of the wavespeed on wavelength. For example, with 17 waves in the annulus the wavespeed is 2.0 mm min^{-1} , while with 7 waves it is 2.3 mm min^{-1} . Thus waves with a wider gap in front of them travel faster, an effect also found in previous studies on chemical waves^{10,24}. Our experiments show that regular rotating chemical structures (clockwise as well as counterclockwise) are stable states of the chemical system if the number of waves N falls within a certain interval, $6 \leq N \leq 25$. These chemical pinwheels were maintained without change for 4–5 days; only the gradual shrinking of the gel prevented maintaining them indefinitely. If there were fewer than 6 waves in the annulus, spontaneous pacemakers appeared from time to time between the waves. If the number of waves exceeded about 25, then interaction between the waves destroyed the regular structure.

Our studies furnish the first experimental evidence that nonlinear chemical systems in circular geometries can indeed show the predicted rotating wave structures²⁵. Also, the realization of an open reactor in which stable chemical patterns can be maintained should open up new experimental possibilities for

the study of spatial self-organization in nonlinear systems. In particular, it will make possible a more rigorous comparison of experiments with recent theoretical results (J. Boissonade, personal communication) based on reaction–diffusion equations.

We thank Mary Gerencser, Paul Moke, Bruce Banner, and Kenneth Pope for their assistance. This research was conducted with the support of the US Department of Energy Office of Basic Energy Sciences and the BP Venture Research Unit.

Received 14 May 1987; accepted 3 August 1987.

1. Turing, A. M. *Phil. Trans. R. Soc. B* **237**, 37–72 (1952).
2. Nicolis, G. & Prigogine, I. *Self-Organization in Nonequilibrium Systems* (Wiley, New York, 1977).
3. Haken, H. *Synergetics* (Springer, Berlin, 1978).
4. Field, R. J. & Burger, M. (eds) *Oscillations and Traveling Waves in Chemical Systems* (Wiley, New York, 1985).
5. Zaikin, A. N. & Zhabotinsky, A. M. *Nature* **225**, 535–537 (1970).
6. Winfree, A. T. *Science* **175**, 634–636 (1972).
7. Smoes, M. L. *J. chem. Phys.* **71**, 4669–4679 (1979).
8. Saul, A. & Showalter, K. in *Oscillations and Traveling Waves in Chemical Systems* (eds Field, R. J. & Burger, M.) Ch. 11 (Wiley, New York, 1985).
9. Agladze, K. I. & Krinsky, V. I. *Nature* **296**, 424–426 (1982).
10. Sevcikova, H. & Marek, M. *Physica* **9D**, 140–156 (1983).
11. Welsh, B. J., Gomati, J. & Burgess, A. E. *Nature* **304**, 611–614 (1983).
12. Agladze, K. I., Krinsky, V. I. & Pertsov, A. M. *Nature* **308**, 834–835 (1984).
13. Wood, P. M. & Ross, J. *J. chem. Phys.* **82**, 1924–1936 (1985).
14. Bodet, J. M., Ross, J. & Vidal, C. *J. chem. Phys.* **86**, 4418–4424 (1987).
15. Müller, S. C., Plesser, T. & Hess, B. *Physica* **24D**, 71–86 (1987).
16. Orban, M. *J. Am. chem. Soc.* **102**, 4311–4314 (1980).
17. Showalter, K. *J. chem. Phys.* **73**, 3735–3742 (1980).
18. Avnir, D. & Kagan, M. *Nature* **307**, 717–721 (1984).
19. Nagypal, I., Bazsa, G. & Epstein, I. R. *J. Am. chem. Soc.* **108**, 3635–3640 (1986).
20. Micheau, J. C., Gimenez, M., Borckmans, P. & Dewel, G. *Nature* **305**, 43–45 (1983).
21. Ouyang, Q. *et al. J. phys. Chem.* **91**, 2181–2184 (1987).
22. Winfree, A. T. in *Oscillations and Traveling Waves in Chemical Systems* (eds Field, R. J. & Burger, M.) Ch. 12 (Wiley, New York, 1985).
23. Winfree, A. T. in *Theoretical Chemistry* Vol. 4 (eds Eyring, H. & Henderson, D.) Vol. 4, 1–51 (Academic, New York, 1978).
24. Showalter, K., Noyes, R. N. & Turner, H. *J. Am. chem. Soc.* **101**, 7463–7469 (1979).
25. Nicolis, G. & Prigogine, I. *Self-Organization in Nonequilibrium Systems* 153–156 (Wiley, New York, 1977).

Role of sea floor organisms in oxygen consumption in the deep North Pacific Ocean

Richard A. Jahnke* & George A. Jackson

Scripps Institution of Oceanography, La Jolla, California 92093, USA

A fundamental challenge facing oceanographers is that of developing a quantitative understanding of deep-ocean recycling of biogenic materials. Such an understanding would not only provide unique insight into the world's carbon cycle and the ultimate fate of anthropogenic carbon dioxide but would also provide powerful constraints for ocean circulation models. Here we present estimates of sedimentary oxygen consumption made using benthic flux and pore-water measurements extrapolated throughout the North Pacific basin as a function of water depth. Below 3 km, sea-floor respiration is a large and perhaps dominant contributor to deep-ocean metabolism. This benthic activity appears to be highly concentrated in regions adjacent to continental margins exhibiting coastal upwelling. More extensive benthic respiration measurements are needed to verify this, especially in the northern and western Pacific Ocean.

Evaluation of deep-ocean metabolism has historically proceeded by modelling the water-column distributions of the major biogenic elements¹⁻⁴. Recent studies have used sediment-trap results to estimate the destruction rate of the sinking biogenic debris⁵⁻⁷. Sinking particles always appear to be degraded predominantly (>70%) in the upper 1,000 m of the water column. Sediment-trap studies have observed only a small (less than a factor of two) decrease in the carbon flux in deeper water between 1,000 m and 4,000 m⁸⁻¹², implying that the extent of degradation of falling particles in the deep (>1,000 m) water column is relatively small. As a result, benthic oxygen consumption and nutrient regeneration may exert a major influence on the concentration of these substances in the water column.

Here we estimate the contribution of the benthos to total deep North Pacific metabolism by first estimating benthic respiration throughout the basin as a function of bottom depth and general location, calculating the total sea-floor oxygen consumption rate as a function of bottom depth, and then dividing this benthic respiration per unit depth interval by the cross-sectional ocean

area at that depth to derive the equivalent oxygen consumption rate per unit volume contributed by the benthos.

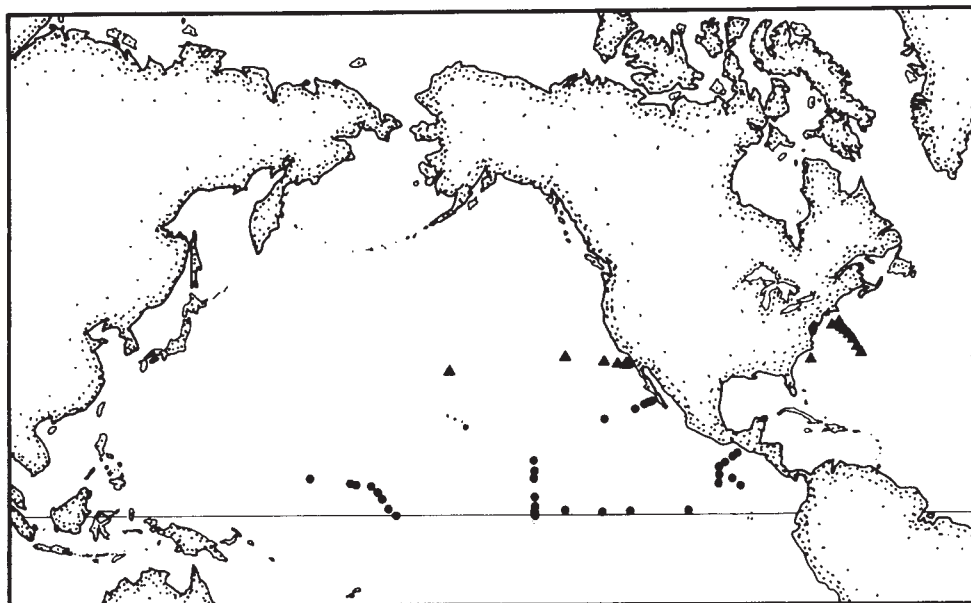
Locations of benthic-flux estimates in the North Pacific from *in situ* benthic chamber experiments¹³ or pore-water nitrate or oxygen profiles (refs 14-17 and P. Froelich, personal communication) are shown in Fig. 1. To enable comparison between different measurements, all values are presented on an oxygen equivalent basis using Redfield stoichiometry ($O_2:C:N:P=138:106:16:1$). (The use of a different stoichiometry such as that proposed by Takahashi *et al.*¹⁸ would not alter the conclusions.) To facilitate the application of these results to large areas of the sea floor, we have grouped the reported fluxes into regions representing areas adjacent to upwelling continental margins, areas adjacent to non-upwelling continental margins, and the equatorial region.

Benthic oxygen fluxes (F_{O_2}) along the east-west transect near the upwelling margin off California¹⁹ are expressed, for convenience, as a function of water depth (Fig. 2). The large value at 3,800 m is the mean of 19 separate benthic chamber measurements¹⁹ and is similar to fluxes estimated from oxygen microelectrode measurements there¹⁴. Although surface-water productivity in the region off California is high, it is not higher than that estimated for the Costa Rica Dome or the Peru upwelling coast. Furthermore, because this station lies to the west of the California borderland, organic input is not significantly augmented by terrigenous or continental shelf materials. Finally, studies adjacent to the Mexican¹⁵, Guatemalan¹⁵ and Panamanian (R. Aller, personal communication) continental margins reveal similar pore-water conditions. We assume that the large fluxes observed along the deep continental rise adjacent to the coastal upwelling region off California are representative of the deep continental slope areas in the eastern Pacific (equation *b* of Fig. 2 legend).

We know of no modern data near the western non-upwelling margin; to estimate benthic fluxes in this region, we rely on benthic chamber measurements obtained in the western Atlantic¹³ (Figs 1, 2). Hinga²⁰ argued that surface productivity is higher in the North Pacific than in the Atlantic, but his Pacific data were mostly from the eastern, upwelling margin whereas his Atlantic data were from the western, non-upwelling margin. Much of the difference that he observed could result from east-west differences. If Hinga is correct, however, our extrapolation underestimates benthic flux in the western Pacific region. Because the fluxes have been extrapolated based on water depth and the western Pacific is significantly deeper than the western Atlantic, especially near the continental margin, it is again likely

* Present address: Skidaway Institute of Oceanography, Savannah, Georgia 31416, USA.

Fig. 1 Station locations at which benthic fluxes have been estimated in the North Pacific^{13-17,19,21}. Pore-water studies are represented by dots, *in situ* benthic chamber measurements by triangles.



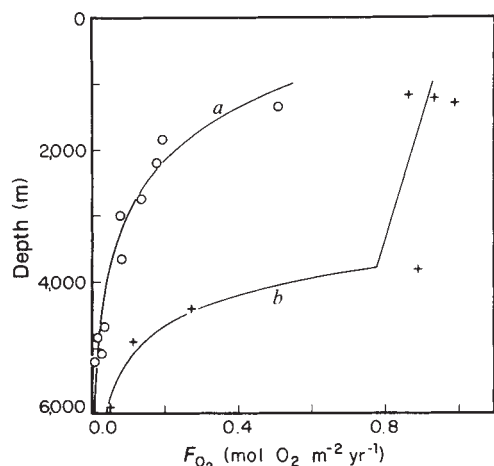


Fig. 2 Benthic oxygen consumption per unit area, F_{O_2} , as a function of water depth, z , for respirometer stations adjacent to (a), the non-upwelling Atlantic and (b), upwelling Pacific coastal margins (from ref. 14). The equations fit to the data are (with F_{O_2} in units of $\text{mol m}^{-2} \text{yr}^{-1}$ and z in m): a, $F_{O_2} = 1.299 e^{-8.5746 \times 10^{-5} z}$ for $1,000 \leq z \leq 6,000$; b, $F_{O_2} = 0.988 - (5.524 \times 10^{-5})z$ for $1,000 \leq z < 3,800$ and $F_{O_2} = 3.153 \times 10^{-23} z^{-6.5947}$ for $3,800 \leq z \leq 6,000$.

that the western Pacific fluxes have been underestimated. The same is true for high-latitude regions. Organic carbon flux to the sea floor at station P (140°W , 50°N) has been estimated from long-term sediment-trap collections¹⁰ to be $0.2 \text{ mol C m}^{-2} \text{yr}^{-1}$. If 90% of this carbon is oxidized, benthic oxygen flux is $\sim 0.23 \text{ mol m}^{-2} \text{yr}^{-1}$. The relationship used in this study yields a value of $0.05 \text{ mol m}^{-2} \text{yr}^{-1}$. Thus, there is good reason to believe that our results underestimate benthic fluxes from these regions and their effects on oceanic chemistry.

Pore-water studies performed near the equator (refs 15–17, 21 and P. Froelich, personal communication) suggest that this region may also be an important contributor to deep sea-floor metabolism. We have assigned fluxes in the centre of the basin based on pore-water diffusion calculations. Because the steep surficial pore-water gradients observed in the eastern Pacific are not sufficiently resolved to calculate the diffusive flux precisely, benthic fluxes in this area are assumed to be equal to those measured by Smith¹⁹ off California. These values are in the middle of the range estimated from pore-water studies¹⁵.

Based on these descriptions of sea-floor fluxes, we estimate the total sediment oxygen consumption (T_{O_2}) for the North

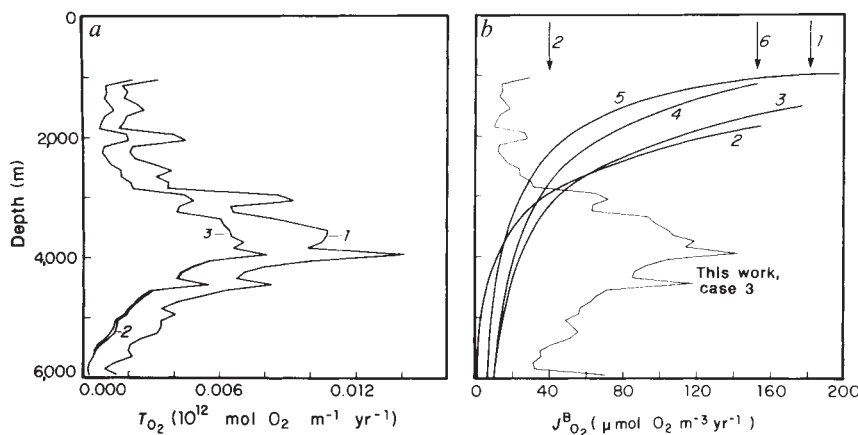
Pacific Ocean as a function of depth and region for three cases as described in the Fig. 3 legend. We use the US Naval Oceanographic Office 5' Digital Bathymetric Data Base to calculate bottom depth at each 5' square. We consider only water depths greater than 1,000 m to focus on deep-ocean cycling. The North Pacific is well suited for this calculation because the nearly horizontal constant-density surfaces²² below 1,000 m preclude benthic regeneration in shallow areas from affecting concentrations in the deep water column by isopycnal mixing.

There is a large maximum in T_{O_2} between 3,000 and 4,000 m (Fig. 3a); this depth range is where there is maximum exposed sea-floor surface area²³. Despite the decrease in benthic fluxes with increasing water depth (Fig. 2), total sea-floor recycling increases in this depth interval because of a larger increase in bottom area. For comparison, the total integrated amount of the organic carbon oxidized in the sediments below 1 km water depth estimated for case 3 is 2% of the ^{14}C primary productivity²⁴, a value consistent with the global estimate of ref. 13.

To compare the benthic recycling rate with other estimates of deep-water metabolism, we have calculated the average (benthic) oxygen consumption rate per unit volume, $J_{O_2}^B$, by dividing T_{O_2} by the oceanic cross-sectional area: $J_{O_2}^B = T_{O_2}/A$ (Fig. 3b). These values represent the amount of oxygen consumption in each cubic metre of sea water which could be attributed to benthic respiration and subsequent lateral exchange. The effect of depth on $J_{O_2}^B$ is similar to that for T_{O_2} (Fig. 3a), with the mid-depth maximum shifted slightly deeper because of the decreased oceanic cross-section at increased depth.

Because it is difficult to make direct measurements of water-column oxygen consumption, most estimates result from fits of chemical tracer distributions to physical-chemical models which do not differentiate benthic respiration. Craig³ used a one-dimensional vertical advection-diffusion-reaction model to estimate the total oxygen consumption rate, $J_{O_2} = 180 \mu\text{mol m}^{-3} \text{yr}^{-1}$ for an upwelling velocity of 6 m yr^{-1} . Using a three-dimensional model of the Pacific with a flat bottom, no benthic respiration and an upwelling velocity of 3 m yr^{-1} , Fiadiero and Craig⁴ estimated J_{O_2} to be $40 \mu\text{mol m}^{-3} \text{yr}^{-1}$ in the deep waters. They also report a better fit to the data with oxygen consumption related to depth by an exponential relationship, given by $J_{O_2} = 480 e^{-1.2(z-1)} \mu\text{mol m}^{-3} \text{yr}^{-1}$ for z , deeper than 1 km, given in km. Hinga²⁰ used the relationship between $\Delta^{14}\text{C}$ and oxygen in deep Pacific waters to estimate J_{O_2} as $151 \mu\text{mol m}^{-3} \text{yr}^{-1}$. Jenkins^{25–27} has used ^3H and ^3He tracers in the upper 1,800 m of the Sargasso Sea in the North Atlantic to study J_{O_2} and the utility of one-dimensional models. He concluded²⁶ that vertical advective-diffusive models are inadequate for describing the distribution

Fig. 3 a, Total benthic oxygen consumption, T_{O_2} , over the North Pacific as a function of depth for three cases. In case 1 the benthic flux is calculated by using $T_{O_2} = F_{O_2}(dA/dz)$, where A is the cross-sectional area of the North Pacific at a given depth z , assuming that the F_{O_2} -depth relationship observed along the upwelling Pacific transect is applicable everywhere. Case 2 uses the upwelling Pacific transect to represent all regions east of 150°W and the Atlantic, non-upwelling transect to represent the area west of 150°W . Case 3 is case 2 modified by assigning benthic fluxes to the region south of 20°N . In the central equatorial region, the fluxes are based on pore-water diffusion calculations. Models of surface pore-water gradients in the eastern region¹⁶ predict a wide range of possible benthic oxygen fluxes, 0.13 – $1.2 \text{ mol O}_2 \text{ m}^{-2} \text{yr}^{-1}$. The easternmost equatorial region is assigned a sedimentary oxygen consumption rate of $0.8 \text{ mol O}_2 \text{ m}^{-2} \text{yr}^{-1}$ because of its similarity to the California transect, where direct benthic flux measurements are available. b, Water-column oxygen consumption rate, $J_{O_2}^B$, which could be attributed to benthic recycling and lateral transport for case 3. Previous estimates based on hydrographic or sediment-trap studies (numbered 1–6) are from refs 3, 4, 6, 5, 7 and 20, respectively. Sediment-trap model fluxes were calculated assuming a mean basin-wide surface-water primary productivity of $100 \text{ g C m}^{-2} \text{yr}^{-1}$.



Previous estimates based on hydrographic or sediment-trap studies (numbered 1–6) are from refs 3, 4, 6, 5, 7 and 20, respectively. Sediment-trap model fluxes were calculated assuming a mean basin-wide surface-water primary productivity of $100 \text{ g C m}^{-2} \text{yr}^{-1}$.

of chemical tracers and that lateral mixing influences their vertical distributions. Jenkins²⁶ found that J_{O_2} between the surface and 1,800 m decreased exponentially with a decay length of ~ 350 m. J_{O_2} at 1,800 m was $187 \mu\text{mol m}^{-3} \text{yr}^{-1}$. These results for the Atlantic are similar to those for the Pacific determined by other methods. Models of sediment-trap results have also yielded estimates of J_{O_2} (Fig. 3b). A comparison of our results with these reported estimates of deep-ocean metabolism indicates that below 3 km, benthic recycling may account for a large portion of the total deep metabolism as estimated by Craig³ and Hinga²⁰, and clearly dominates over the other estimates (Fig. 3b). This conclusion is consistent with the results of deep-ocean sediment-trap studies⁸⁻¹², despite the fact that we are probably underestimating the fluxes in the western and northern regions.

Based on these calculations, we conclude that recently published mathematical representations of particle fluxes in the ocean⁵⁻⁷ overestimate the extent of water column mineralization in the deep ocean. Although the majority of the biogenic flux out of the photic zone is recycled in the upper 500–1,000 m, most of the material escaping this active upper region appears to survive to the sea floor. The fate of biogenic particles sinking through the oceanic water column may be better represented by a two-layer system where very active recycling is restricted to the upper 500–1,000 m. Below this depth, destruction of the raining particles proceeds at a very slow rate, with most of the particles reaching the sea floor. This simplified two-layer description of the ocean better represents available data but will undoubtedly be modified as more observations are made.

Our calculation implies that recycling of biogenic debris at the sea floor is significant and possibly dominant in deep-ocean mineralization. This recycling increases dramatically towards the eastern boundary. The calculation presented here suggests that perhaps as much as 50% of all of the deep benthic oxygen consumption occurs within 500 km of the ocean boundary. Thus, the continental margin regions may significantly influence the composition of the ocean's interior. Physical models that attempt to describe water-mass movements and non-conservative biogenic tracer distributions may better represent deep-ocean respiration as a reaction term at the ocean boundaries, with reasonably conservative behaviour within the ocean's interior.

Further refinements in the description of deep Pacific Ocean metabolism will require additional determinations of benthic respiration rates, especially in the northern and western regions. Although this discussion has concerned only the North Pacific, we strongly expect the generalizations to be applicable to other major ocean basins.

We thank E. Stewart for help in manipulating the data, S. Smith for providing the benthic depth data, and J. Reid and S. Anderson for providing hydrographic data. This work was funded by the US Department of Energy.

Received 10 February; accepted 28 July 1987.

1. Fiadiero, M. thesis, Univ. California/Scripps Inst. Oceanogr. (1975).
2. Craig, H. *J. geophys. Res.* **74**, 5491–5506 (1969).
3. Craig, H. *J. geophys. Res.* **76**, 5078–5086 (1971).
4. Fiadiero, M. & Craig, H. *J. mar. Res.* **36**, 323–355 (1978).
5. Suess, E. *Nature* **288**, 260–263 (1980).
6. Betzer, P. R. *et al. Deep Sea Res.* **31**, 1–11 (1984).
7. Pace, M. L., Knauer, G. A., Karl, D. M. & Martin, J. M. *Eos* **67**, 1036 (1986).
8. Honjo, S. *Proc. Global Ocean Flux Study Workshop* 306–324 (National Academy of Sciences, Woods Hole, Massachusetts, September 1984).
9. Fischer, K. M. thesis, Oregon State Univ. (1984).
10. Asper, V. L. thesis, Massachusetts Inst. Technol./Woods Hole Oceanogr. Inst. (1986).
11. Honjo, S., Manganini, S. J. & Cole, J. J. *Deep Sea Res.* **29**, 609–625 (1982).
12. Lorenzen, C. J., Welschmeyer, N. A. & Copping, A. E. *Deep Sea Res.* **30**, 639–643 (1983).
13. Smith, K. L. Jr & Hinga, K. R. *The Sea* Vol. 8 (ed. Rowe, G. T.) 331–370 (Wiley, New York, 1983).
14. Reimers, C. E. & Smith, K. L. Jr *Limnol. Oceanogr.* **31**, 305–318 (1986).
15. Jahnke, R. A., Emerson, S. R. & Murray, J. W. *Limnol. Oceanogr.* **27**, 610–623 (1982).
16. Grundmanis, V. & Murray, J. W. *Geochim. cosmochim. Acta* **46**, 1101–1120 (1982).
17. Jahnke, R., Heegle, D., Emerson, S. & Grundmanis, V. *Earth planet. Sci. Lett.* **61**, 233–256 (1982).
18. Takahashi, T., Broecker, W. S. & Langer, S. J. *geophys. Res.* **90**, 6907–6924 (1985).
19. Smith, K. L. Jr *Limnol. Oceanogr.* **32**, 201–220 (1987).
20. Hinga, K. R. *Deep Sea Res.* **32**, 117–126 (1985).

21. Emerson, S. *et al. Earth planet. Sci. Lett.* **49**, 57–80 (1980).
22. Craig, H., Broecker, W. S. & Spencer, D. *GEOSECS Pacific Expeditions* Vol. 4 (National Science Foundation, Washington, 1981).
23. Menard, H. W. & Smith, S. M. *J. geophys. Res.* **71**, 4305–4325 (1966).
24. Eppley, R. W. & Peterson, B. J. *Nature* **282**, 677–680 (1979).
25. Jenkins, W. J. *Nature* **300**, 246–248 (1982).
26. Jenkins, W. J. *J. mar. Res.* **38**, 533–569 (1980).
27. Jenkins, W. J. *J. mar. Res.* **40**, Suppl. 265–290 (1982).

Geochemical formation of organosulphur compounds (thiols) by addition of H_2S to sedimentary organic matter

Appathurai Vairavamurthy & Kenneth Mopper

Division of Marine and Atmospheric Chemistry,
Rosenstiel School of Marine and Atmospheric Science,
University of Miami, 4600 Rickenbacker Causeway,
Miami, Florida 33149-1098, USA

Organosulphur compounds constitute a significant fraction of organic matter in both recent and ancient iron-poor marine sediments as well as in petroleum generated in such sediments^{1,2}. Most studies on organosulphur compounds in marine sediments have focused on hydrophobic compounds, for example, thiophenes³⁻⁶, sulphur-containing hopanoids⁷ and volatile compounds. Recently, hydrophilic, organic thiols were found in significant concentrations in anoxic marine sediments^{8,9}. Several studies¹⁰⁻¹³ have shown that H_2S formed from bacterial sulphate reduction can become incorporated into organic matter during early diagenesis, however, the exact chemical mechanisms by which this sulphur reacts with sedimentary organic matter are not well understood. Here we describe the widespread occurrence of 3-mercaptopropionic acid in coastal marine sediments (Biscayne Bay, Florida) and present evidence for its formation by an abiotic reaction between hydrogen sulphide and acrylic acid, which is a cleavage product of the common algal osmolyte β -dimethylsulphoniopropionate^{14,15}. The reaction probably occurs by nucleophilic addition of bisulphide (HS^-) ion to the activated double bond in the α,β -unsaturated carbonyl system of acrylic acid. Reactions between hydrogen sulphide and activated unsaturated bonds in organic molecules are probably responsible for the formation of other thiols in sediments and could represent a major pathway for the incorporation of sulphur into sedimentary organic matter during early diagenesis.

Sediment samples were collected using a plastic corer, in the intertidal zones of Biscayne Bay, Florida over seven months covering summer and winter seasons. The cores were subsampled in an anoxic glove box and pore-water samples were obtained by centrifugation. Some sediments and pore-water samples were treated with tributylphosphine, which releases bound thiols by cleavage of the sulphur-sulphur bonds¹⁶. Thiols were determined by pre-column fluorescence derivatization and high-performance liquid chromatography¹⁷.

More than 20 thiols were detected in sediment pore-water samples and, in nearly all samples, 3-mercaptopropionic acid (3-MPA) was found as a major thiol (Fig. 1). Methane thiol and glutathione were also present in significant concentrations. The thiols were identified by co-injection with authentic compounds under different chromatographic and derivatization conditions⁹ as well as by using another fluorescence derivatization method with monobromobimane¹⁸. 3-MPA was present at nanomolar to micromolar levels, with maximum concentrations in the summer during periods of high sulphate reduction activity¹⁹.

Thiols play a variety of biochemical roles as coenzymes, in the maintenance of cytoplasmic redox states, binding metals in the active sites of enzymes and electron-transport components, and in the capture and detoxification of metals. Thus, a biochemical origin can be invoked for many of the thiols present

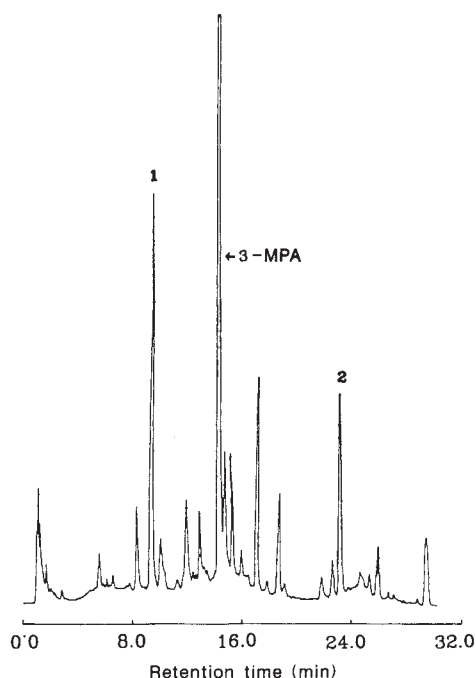


Fig. 1 A typical chromatogram of thiols in a sediment pore-water sample (Biscayne Bay, Florida) using high-performance liquid chromatography pre-column *o*-phthalaldehyde derivatization and fluorescence detection. Peaks numbered 1 and 2 correspond to glutathione and methane thiol respectively.

in sediments (for example methane thiol, glutathione and coenzyme M). But a biological origin is not known for 3-MPA. An abiotic formation can be suggested based on the reaction between hydrogen sulphide and acrylic acid by the Michael addition mechanism^{20,21}. Acrylic acid is derived in marine sediments from the cleavage of β -dimethylsulphoniopropionate, which occurs in many algae and seagrasses as an osmolyte^{14,15}. Hydrogen sulphide formed by bacterial dissimilatory sulphate reduction is probably present in concentrations in excess of acrylic acid in most strongly reducing marine sediments and salt marshes. Typical pore-water concentrations of hydrogen sulphide in anoxic Biscayne Bay sediments are 1–5 mM.

We carried out kinetic studies on the abiotic formation of 3-MPA using excess bisulphide and limiting acrylic acid concentrations. The reactions were performed in amber glass bottles using de-aerated, nitrogen sparged sea water (pH 8.3 ± 0.1 and salinity 35). Concentrated stock solutions of reagent grade sodium hydrogen sulphide (Morton Thiokol) and acrylic acid (Aldrich) were made fresh in filtered, de-aerated sea water and added to reaction bottles in appropriate amounts to result in 5 or 10 mM of bisulphide and 50–500 μ M of acrylic acid. The addition reaction, which is bimolecular with respect to the reactants, should follow a pseudo first-order reaction kinetics with an excess of one reactant. As shown in Fig. 2, the reaction does have pseudo first-order reaction kinetics with excess sulphide concentrations. The effect of temperature on the rate of 3-MPA formation in sea water was also investigated. Calculations using the Arrhenius equation indicate an activation energy of 15.0 kcal per mole for the addition reaction. Estimates of the rate of 3-MPA formation at environmentally relevant temperatures, such as 35 °C and 15 °C from the Arrhenius plot (Fig. 3) indicate pseudo first-order rate constants of 9.1×10^{-3} per day and 1.7×10^{-3} per day respectively for reaction with excess sulphide (10 mM). Therefore, in most anoxic marine sediments with hydrogen sulphide concentrations of 1–10 mM, rate constants of the same order of magnitude can be expected for the

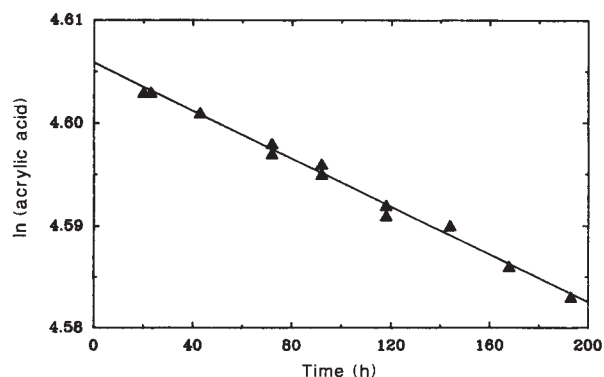


Fig. 2 Pseudo first-order reaction kinetics for the formation of 3-MPA with excess bisulphide (5 mM) and limiting acrylic acid concentrations. The reaction was carried out in de-aerated, filter-sterilized sea water (salinity 35 and pH 8.4) at 23 °C. Acrylic acid concentrations were calculated by subtracting the amount reacted to form 3-MPA from the starting concentration. The straight line had an equation; $y = -0.00011x + 4.6058$ and $r^2 = 0.993$ (r^2 is the coefficient of determination).

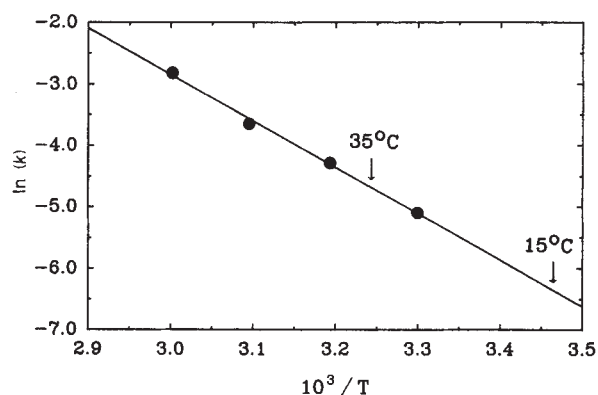


Fig. 3 Arrhenius plot illustrating temperature dependence of 3-MPA formation. The constant, k , is the pseudo first-order rate constant for 3-MPA formation with excess sulphide (10 mM) and limiting acrylic acid concentrations. T is the absolute temperature.

formation of 3-MPA. The estimated values agree with the seasonal variation of 3-MPA observed in sediments, although factors other than temperature may also be involved.

Formation of 3-MPA by Michael addition involves attack of the nucleophile, bisulphide ion (HS^-), at the β -carbon in acrylic acid^{20,21}. It is also possible that hydrogen sulphide reacts with acrylic acid by an electrophilic mechanism that involves an initial protonation step followed by HS^- attack on the double bond²⁰. But strong electron-withdrawing groups (such as $-\text{C}=\text{O}$, $-\text{COOH}$, $-\text{COOR}$ and $-\text{CN}$) attached to the carbon-carbon double bond, as in acrylic acid, tend to lower its reactivity towards electrophiles thereby favouring the nucleophilic mechanism. It has been shown by Gershbein and Hurd²² that base catalysis significantly increases the reactivity of hydrogen sulphide with acrylonitrile and methyl acrylate in non-aqueous reaction media which supports the nucleophilic addition. Although the reactions studied here were carried out in an aqueous medium, the addition mechanism is presumably analogous. If the attack of HS^- had occurred at the α -carbon of acrylic acid, the resulting compound would have been 2-MPA. But our experiments have not shown the formation of this thiol. Addition of HS^- at the β -carbon leading to the formation of 3-MPA is favoured by the presence of the conjugated, α,β -unsaturated carbonyl system in acrylic acid which permits formation of a resonance-stabilized anion intermediate^{20,21}.

Several analogues of acrylic acid with α,β -unsaturated carbonyl systems (for instance, crotonic acid, methacrylic acid, fumaric acid and maleic acid) also reacted with H_2S in sea water medium yielding the corresponding thiols. These reactions were carried out using conditions similar to those used for studying the addition of hydrogen sulphide to acrylic acid. Thiomalic acid, presumably formed from reaction of fumaric acid with bisulphide, was tentatively identified in a few sediments. The rate of abiotic formation of 3-MPA was found to increase with increase in salinity of the medium at neutral and basic pH values. Acrylic acid exists as carboxylate anion at these pH values and increase in salinity probably induces ion-pair formation resulting in partial neutralization of the negatively charged carboxyl group. This effect could enhance the electrophilicity of the β -carbon in acrylic acid and its ability to react with the bisulphide ion. In support of this explanation, addition of bisulphide to α,β -unsaturated esters and nitriles (such as methyl acrylate and acrylonitrile) occurs at significantly higher rates than α,β -unsaturated acids in aqueous medium (Table 1).

Table 1 Pseudo first-order rate constants for Michael addition of HS^- (5mM) to different α,β -unsaturated compounds

α,β -unsaturated compound	k (day $^{-1}$) at 23 °C
Acrylic acid	2.6×10^{-3}
Fumaric acid	6.5×10^{-3}
Methyl acrylate	6.4×10^{-2}
Acrylonitrile	1.4×10^{-2}

Thiols formed in reactions were detected by high-performance liquid chromatography¹⁷; identification by co-injection with authentic standards was made only for 3-MPA and thiomalic acid formed with acrylic acid and fumaric acid respectively. Thiols formed from methyl acrylate and acrylonitrile were quantified based on the response factor for 3-MPA.

The presence and high concentrations of the unusual thiol, 3-MPA in reducing marine sediments can be readily explained in terms of a simple nucleophilic addition between hydrogen sulphide and acrylic acid. Similar abiotic reactions involving hydrogen sulphide, as well as polysulphides, may also contribute significantly to the formation of other organosulphur compounds in sediments^{10-12,23}. Recently, Brassel *et al.*³ speculated that reaction between hydrogen sulphide and phytadienes gives rise to isoprenoid thiophenes, which occur widely in recent and ancient marine sediments. Our results suggest that abiotic nucleophilic addition of hydrogen sulphide, and possibly polysulphides, to activated unsaturated bonds of organic molecules represents a major chemical pathway for the incorporation of sulphur into sedimentary organic matter during early stages of diagenesis.

We thank K. Van Scoy for assistance with field data and B. F. Taylor, G. R. Harvey and R. Kiene for helpful comments. This material is based upon work supported by the NSF.

Received 4 May; accepted 21 August 1987.

1. Tissot, B. P. & Welte, D. H. *Petroleum Formation and Occurrence* (Springer, Berlin, 1984).
2. Gransch, J. A. & Posthuma, J. in *Advances in Organic Geochemistry 1973* (eds Tissot B. & Biener F.) 727-739 (Technip, Paris, 1974).
3. Brassel, S. C., Lewis, C. A., de Leeuw, J. W., de Lange, F. & Sinninghe Damste, J. S. *Nature* **320**, 160-162 (1986).
4. Sinninghe Damste, J. S., ten Haven, H. L., de Leeuw, J. W. & Schenck, P. A. *Org. Geochem.* **10**, 791-805 (1986).
5. Sinninghe Damste, J. S., Kock-van Dalen, A. C., de Leeuw, J. W. & Schenck, P. A. *Tetrahedron Lett.* **28**, 957-960 (1987).
6. Sinninghe Damste, J. S. *et al.* **51**, (in the press).
7. Valisollalao, J., Perakis, N., Chappe, B. & Albrecht, P. *Tetrahedron Lett.* **25**, 1183-1186 (1984).
8. Luther, G. W. III, Church, T. M., Scudlark, J. R. & Cosman, M. *Science* **232**, 746-749 (1986).
9. Mopper, K. & Taylor, B. F. in *Organic Marine Chemistry ACS Symposium Series 305* (ed. Sohn, M. L.) 324-339 (Am. Chem. Soc., Washington DC, 1986).
10. Sinninghe, J. S. & de Leeuw, J. W. *Int. J. Environ. Anal. Chem.* **28**, 1-19 (1987).
11. Francois, R. *Geochim. cosmochim. Acta* **51**, 17-27 (1987).
12. Aizenshtat, Z., Stoler, A., Cohen, Y. & Nielsen, H. in *Advances in Organic Geochemistry 1981* (eds Bjorøy, M. *et al.*) 279-288 (Wiley, Chichester, 1983).

13. Casagrande, D. J. *et al. Nature* **282**, 559-600 (1979).
14. Charlson, R. J., Lovelock, J. E., Andreae, M. O. & Warren, S. G. *Nature* **326**, 655-661 (1987).
15. Vairavamurthy, A., Andreae, M. O. & Iverson, R. L. *Limnol. Oceanogr.* **30**, 59-70 (1985).
16. Ruegg, U. Th. & Rudinger, J. in *Methods of Enzymology XLVII* (eds Hirs, C. H. W. & Timasheff, S. N.) 111-116 (Academic, New York, 1977).
17. Mopper, K. & Delmas, D. *Analyt. Chem.* **56**, 2558-2560 (1984).
18. Newton, G. L., Dorian, R. & Fahey, R. C. *Analyt. Biochem.* **114**, 383-387 (1981).
19. Jørgensen, B. B. *Limnol. Oceanogr.* **22**, 814-832 (1977).
20. Morrison, R. T. & Boyd, R. N. *Organic Chemistry* (Allyn and Bacon, Massachusetts, 1987).
21. Bergmann, E. D., Ginsburg, D. & Pappo, R. in *Organic Reactions Vol. 10* (eds Adams, R. *et al.*) 179-555 (Wiley, New York, 1959).
22. Gershbein, L. L. & Hurd, C. D. *J. Am. chem. Soc.* **69**, 241-242 (1947).
23. Dinur, D., Spiro, B. & Aizenshtat, Z. *Chem. Geol.* **31**, 37-51 (1980).

Dental development of the Taung skull from computerized tomography

Glenn C. Conroy* & Michael W. Vannier†

* Department of Anatomy and Neurobiology/Anthropology

† Mallinckrodt Institute of Radiology, Washington University School of Medicine, St Louis, Missouri 63110, USA

Just over 60 years ago, Dart's description and analysis of the Taung child's skull¹ triggered an intellectual revolution about human origins. Recently, several authors have suggested that one of the most significant hominid-like traits of australopithecines, delayed maturation, may not after all be valid²⁻⁴. This is a radical departure from Mann's⁵ classic study of australopithecine maturation and palaeodemography based on dental eruption patterns. The resolution of this debate has important implications for the history of the biological and social evolution of the human species. In view of the controversies generated by recent studies⁶, and particularly because the Taung skull is the type specimen of *Australopithecus africanus*, we have investigated the relevant anatomy of the Taung 'child' using computerized tomography. We conclude that the Taung 'child' shows some important dental maturational affinities with great apes, although as Dart¹ noted, other hominid-like features are clearly present.

Previous radiographic analyses of the Taung skull have been of limited value because of the dense mineralization and calcified matrix in the skull⁷. For this reason, important intracranial structures like the developing permanent dentition and paranasal sinuses have never before been clearly visualized. This is particularly unfortunate in that dental development patterns have been critical in evaluating australopithecine maturational affinities^{1,4-5}.

Recent advances in high-resolution computerized tomography (CT) have made it possible to see previously hidden structures in the Taung skull⁸⁻¹¹. We have recently used CT to scan the entire Taung skull in sagittal, coronal and transaxial planes. These serial 2-mm scans show for the first time, and with high resolution and accuracy, the state of development of all the unerupted permanent dentition and the pneumatized portions of the facial skeleton (Figs 1 and 2). These views allow more refined inferences about the developmental patterns of this important specimen to be drawn.

It is clear that the first permanent molars had recently erupted but were not yet in functional occlusion^{12,13}. The critical question is whether Taung follows a 5-7-year-old human dental growth trajectory^{1,5} or a growth trajectory like that of a 3-4-year-old ape²⁻⁴, or something in between. The answer has implications for the evolutionary history of delayed maturation, a uniquely human characteristic. It was recently suggested^{3,4} that the great-ape-like situation prevailed for australopithecines in general (particularly the 'gracile' australopithecines)—although neither study was able to analyse the Taung dentition because of the obscuring calcified matrix. To compare the Taung skull with living great apes and hominids, a modern human skull and chimpanzee skull, both at the same stage of M1 eruption as the Taung skull, were scanned in 2-mm serial sections.

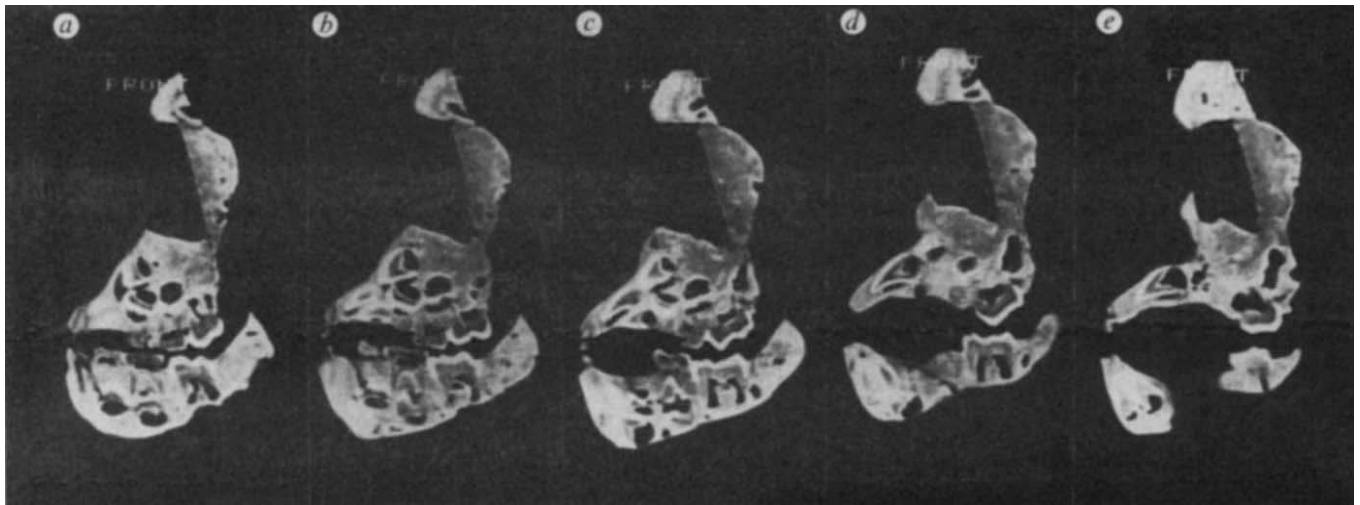


Fig. 1 a-e, Five contiguous 2-mm thick CT sections taken along the left tooth row (lateral to medial) of the Taung facial fragment and mandible showing the complete developing dentition.

Methods. Serial 2-mm thick, high-resolution scans were obtained using a Siemens Somatom DR-3 Scanner at Hillbrow Hospital, Johannesburg. Extended range bone reconstruction was used because the CT density of the mineralized bone, enamel and stone matrix was well beyond 1,000 Hounsfield units, the upper density limit for normal CT scanning.

The developing dentition in the Taung specimen as revealed by CT is as follows (upper dentition): I1, crown formation nearly complete, no root development; I2, crown formation nearly complete, no root development; C, calcification of about three-quarters of the crown completed, no root development; P3, crown formation about three-quarters completed, no root development; P4, crown formation about three-quarters completed, no root development; M1, crown formation completed and approximately three-quarters of root development usually completed (root length, ~7.5 mm); M2, calcification of about one-half of the crown completed, no root development; M3 = no trace of any dental development (none would be expected in great apes or hominids until root formation of the M2 was well under way). Because formation of upper incisors in primates is not normally complete until after that of the lower incisors, it follows that the lower incisors would be slightly advanced in crown formation over the upper incisors. Nonetheless, no root formation can be detected on the lower permanent incisors of the Taung skull. Likewise, the lower permanent canines and premolars appear slightly more advanced in crown development than their counterparts in the upper jaw.

Comparison of these stages of dental development with those of mean chronological calcification times of the developing dentition in modern great apes and hominids^{4,14-19} reveals that, overall, the pattern of the Taung specimen is most like that of a 3.0-4.0-year-old pongid. This is based on the observation that at first molar eruption stage in pongids, none of the other permanent teeth have as yet developed any significant root structure. In contrast, at a similar stage in humans, incisors, canines and third premolars have developed at least some root structure.

This difference between extant great apes and hominids is manifested in the following way. Modern humans are distinguished from great apes in the eruption timing of the first permanent incisors and first permanent molars^{12,14}. Modern humans are unique in that both permanent incisors and first molars erupt during one relatively brief time span in the growth period (at about 6 years of age). In great apes, on the other hand, the eruption of permanent first incisors is delayed for about two years after first molar eruption. Thus, at first molar eruption stage in great apes the permanent incisors are usually still low in their alveolar crypts and the permanent incisors have little, if any, root development. The Taung dentition clearly reflects this condition.

It has previously been suggested that this great-ape-like M1-I1 eruption pattern characterized 'gracile' australopithecines (such as LH-2 and STS 24) whereas the human-like sequence characterized 'robust' australopithecines and some early *Homo* (SK 61, 62, 63, TM 1536, KNM-ER 820) (refs 12 and 20-21, but see ref. 22). Because the developing dentition of the Taung skull had previously been impossible to visualize, its dental (and phylogenetic) affinities were open to speculation². To the extent that these M1-I1 differences between 'gracile' and 'robust' australopithecines^{12,22,23} are valid, we can now state unequivocally that the Taung skull falls into the pattern of the 'gracile' australopithecines.

Note also that the Taung specimen retained the great-ape-like 'retardation' of canine formation (see ref. 14) which suggests (but does not necessarily prove) that canine eruption was delayed as well. The dental eruption sequence in pongids is normally M1 I1 I2 M2 [P3 P4] C M3; the usual pattern in humans is M1 I1 I2 [P3 C P4] M2 M3 (brackets indicate variability in the sequence greater than or equal to 20%)²⁴⁻²⁶. Thus, the canine erupts before the M2 in humans but after the M2 in great apes. The great-ape-like condition is also found in 'gracile' australopithecines, like MLD2 (ref. 13).

One other clear great-ape-like feature of the developing dentition of the Taung skull is the horizontal alignment of the developing permanent incisors. This undoubtedly reflects the greater degree of subnasal alveolar prognathism in the Taung 'child' compared to modern humans. The CT scans reveal other aspects of the Taung skull not previously known. For example, a well developed maxillary sinus is clearly evident (Fig. 2). The volume of the right maxillary sinus, 3.5 cm³, is about the same as in young chimpanzees and humans of equivalent M1 eruption stage even though the overall size of the Taung skull is much smaller. The volume of adult human maxillary sinuses normally ranges between 8 and 15 cm³ and grows about 2 mm in height per year^{27,28}. The maximum height of the maxillary sinus in the Taung skull is about 12 mm; maximum width is about 13 mm, nearly the same as in a 5-6-year-old modern child²⁹. Pneumatization has also extended into the zygoma and hard palate. This is intriguing because an intrapalatal extension of the maxillary sinus has only been reported in chimpanzees and robust australopithecines among higher primates^{30,31}. This extensive pneumatization in such a young skull seems extraordinary.

Previous studies^{1,32} were not able to verify the status of the frontal sinus in the Taung skull. The sagittal CT scans clearly



Fig. 2 A 2-mm thick coronal section through the dp4 showing the large right maxillary sinus and the developing permanent P4 (the left maxillary sinus is matrix-filled). Also note the right nasolacrimal duct and the pneumatization of the zygoma.

show that no frontal sinus was present (Fig. 1). Although such sinuses are present in adult australopithecines³¹, the lack of one in this specimen is not unexpected; the frontal sinus is usually not visible radiographically above nasion in modern children until the age of 6 or 7. A small sphenoid sinus is also present.

It thus appears that the Taung 'child' possesses an interesting mosaic of features. The dental development patterns revealed here by CT are clearly comparable to those of a 3–4-year-old great ape. In addition, the precocious development of the paranasal sinuses, particularly the intrapalatal extensions of the maxillary sinuses, combined with the horizontal alignment of the developing permanent incisors, confirm the retention of some great-ape-like growth mechanism in the Taung facial skeleton². However, these great-ape-like features must be weighed against the undoubted hominid-like features stressed by Dart in his initial description of the skull, which include lack of supra-orbital ridges, shape of the orbits, lack of diastemata in the tooth rows, position of the friamen magnum, and overall brain shape¹. This complex mosaic of craniodental features shows that the Taung 'child' is not a little human, but just as important, it is not a little ape.

We thank the Anatomy staff, University of the Witwatersrand, for their hospitality in Johannesburg. We particularly thank Professor P. V. Tobias for permission to undertake the CT studies on the Taung skull in his care. We also thank the members of the Department of Radiology, Hillbrow Hospital, Johannesburg, for their help with the CT studies: R. Van der Riet, I. Amla, Dr. B. Epstein, B. Mathopa, and S. Ligarabe. We also thank R. Knapp for his assistance and Drs. M. C. Dean, A. D. Beynon, B. H. Smith and A. E. Mann for helpful comments on various aspects of the work. Financial support was provided by the L.S.B. Leakey Foundation, the Council for International Exchange of Scholars, and the NSF.

Received 28 July; accepted 24 August 1987.

1. Dart, R. A. *Nature* **115**, 195–199 (1925).
2. Bromage, T. G. in *Hominid Evolution: Past, Present and Future* (ed. Tobias, P. V.) 239–246 (Liss, New York, 1985).
3. Bromage, T. G. & Dean, M. C. *Nature* **317**, 525–527 (1985).
4. Smith, B. H. *Nature* **323**, 327–330 (1986).

5. Mann, A. E. *Palaeodemographic Aspects of the South African Australopithecines* (University of Pennsylvania, Philadelphia, 1975).
6. Lewin, R. *Science* **235**, 748–750 (1987).
7. Skinner, M. F. & Sperber, G. H. *Atlas of Radiographs of Early Man* (Liss, New York, 1982).
8. Conroy, G. C. & Vannier, M. W. *Science* **226**, 456–458 (1984).
9. Conroy, G. C. & Vannier, M. W. in *Hominid Evolution: Past, Present and Future* (ed. Tobias, P. V.) 419–426 (Liss, New York, 1985).
10. Conroy, G. C. & Vannier, M. W. in *Proc. 10th Congr. Int. Primatol. Soc.* Vol. 1 (eds Else, J. & Lee, P.) 212–222 (Cambridge University Press, 1986).
11. Zonneveld, F. W. & Wind, J. in *Hominid Evolution: Past, Present and Future* (ed. Tobias, P.) 427–436 (Liss, New York, 1985).
12. Dean, M. C. *Am. J. phys. Anthropol.* **67**, 251–258 (1985).
13. Wallace, J. A. thesis Univ. Witwatersrand (1972).
14. Dean, M. C. & Wood, B. A. *Folia Primatol.* **36**, 111–127 (1981).
15. Swindler, D. R. in *Nonhuman Primate Models for Human Growth and Development*, 69–86 (Liss, New York, 1985).
16. Van der Linden, F. P. G. M. & Duterloo, H. S. *Development of the Human Dentition: an Atlas* (Harper and Row, New York, 1976).
17. Gustafson, G. & Koch, G. *Odont. Rev.* **25**, 297–306 (1974).
18. Moorrees, C. F. A., Fanning, E. A. & Hunt, E. E. *J. dent. Res.* **42**, 1490–1502 (1963).
19. Schour, I. & Massler, M. *J. am. dent. Ass.* **28**, 1153–1160 (1941).
20. Broom, R. & Robinson, J. T. *Nature* **167**, 443 (1951).
21. Broom, R. & Robinson, J. T. *Transv. Mus. Mem.* **6**, 1–123 (1952).
22. Grine, F. E. *Am. J. phys. Anthropol.* **72**, 353–360 (1987).
23. Beynon, A. D. & Wood, B. A. *Nature* **326**, 493–496 (1987).
24. Smith, H. B. & Garn, S. M. *Am. J. phys. Anthropol.* (in the press).
25. Schultz, A. H. *Am. J. phys. Anthropol.* **19**, 489–581 (1935).
26. Schultz, A. H. *Contr. Embryol.* **28**, 1–63 (1940).
27. Moss-Salentijn, L. in *Surgery of the Paranasal Sinuses* (eds Blitzer, A., Lawson, W. & Friedman, W. H.) 1–22 (Saunders, Philadelphia, 1985).
28. Ritter, F. N. *The Paranasal Sinuses* (Mosby, St Louis, 1978).
29. Odita, J. C., Akamaguna, A. I., Ogisi, F. O., Amu, O. D. & Ugobodaga, C. I. *Pediatr. Radiol.* **16**, 365–367 (1986).
30. Cave, A. J. E. & Haines, R. W. *J. Anat.* **74**, 493–523 (1940).
31. Tobias, P. V. *Olduvai Gorge* Vol. 2 (Cambridge University Press, 1967).
32. Broom, R. & Schepers, G. W. H. *Transv. Mus. Mem.* **2**, 1–272 (1946).

Optimality, plasticity and selective regime in fig wasp sex ratios

Edward Allen Herre

Smithsonian Tropical Research Institute, PO Box 2072, Balboa, Republic of Panama and Department of Biology, University of Iowa, Iowa City, Iowa 52242, USA

In thirteen species of fig-pollinating wasps the sex ratios of broods with one or two foundress mothers show qualitative agreement with the optimal sex ratios predicted by local mate competition theory. The deviations from these predictions are not random: for a given number of foundresses, the deviations are least in the wasp species that naturally encounter that number of foundresses most frequently. Furthermore, the difference in sex ratio between broods with one and broods with two foundresses is greatest in those species of wasp in which broods of both type commonly occur. These results show the importance of the selective regime in shaping the evolution of sex-ratio adjustment, and suggest that adaptive behavioural plasticity is more developed in species subject to more variable selective regimes.

Sex ratio theory offers one of the clearest opportunities for the study of precision of adaptation because it provides quantitative predictions of the outcome of natural selection of a trait with a large and direct effect on fitness^{1–3}. Hamilton has pointed out⁴ that Fisher's argument⁵ for equal investment in the two sexes depends crucially on the structure of the mating population. He showed that selection for female-biased sex ratios exists when one or a few foundress mothers contribute offspring to isolated mating sub-populations (broods), from which mated females disperse to found new broods. Such conditions characterize the natural history of fig-pollinating wasps. These and other attributes of fig wasps make them useful for studies of sex ratio^{6–11}. For example, in all the species considered in this study, the number of male and female offspring (brood sex ratio) could be associated with the number of foundress mothers contributing to broods within individual fig fruits. Further, the number of foundresses per fig varies so that brood sex ratios can be associ-

ated with a variety of intensities of local mate competition (LMC). Theory predicts increasingly female-biased brood sex ratios with decreasing numbers of foundresses (refs 1, 6–8, 12–16, see Fig. 1). There is also variation among fig wasp species in the level of inbreeding (sib-mating)⁷. Here, theory predicts that more inbred species should exhibit more female-biased broods for any given number of foundresses^{7,8,13,16}. Finally, the current selective regime of discrete LMC situations (the distribution of numbers of foundress mothers per brood) is easily assayed and might reflect longer-term selective history. Therefore, quantitative deviations from theoretical optima for specific LMC situations in different species can be comparatively assessed in the context of their different selective regimes.

For each of 13 species of fig-pollinating wasps from the vicinity of Barro Colorado Island, Republic of Panama, I collected the following data: (1) the distribution of number of foundresses per brood collected between January 1981 and July 1986, a period corresponding to 40 or 50 fig wasp generations; and (2) the number of male and female progeny (brood sex ratio) over the range of numbers of foundresses encountered in the samples for each species. Using the estimates of inbreeding obtained from the data (refs 7, 8, Table 1), I generated predictions of optimal sex ratios for broods with one and two foundresses for each species using the equation

$$p = (1 - m)(2n - 1)/(4n - 1)$$

where p is the optimal brood sex ratio, m is the inverse of the number of foundresses contributing to the brood, and n is the inverse of the proportion of sib-mated females in the species^{7,8,16}. I then calculated the deviations of the means of the empirically determined sex ratios for one and two foundresses from the theoretically predicted sex ratios for one and two foundresses. Finally, I calculated the difference between the means of one- and two-foundress broods for each fig wasp species (see Fig. 1).

I ranked the species in order of both magnitude and direction of deviation of the observed from the expected sex ratio for broods with one and two foundresses and in order of the frequency with which single or multiple foundress cases were actually encountered in the samples. For both the one and two foundress case, the deviations of the actual sex ratios from those predicted by the model are greatest in the fig wasp species which encounter those particular cases the least frequently (Spearman rank correlation coefficient, $R_s = 0.665$, $n = 13$, $P < 0.02$, for

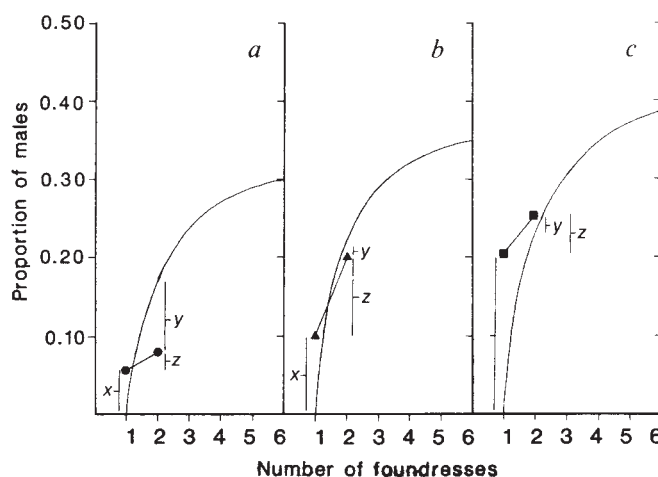


Fig. 1 The theoretically predicted values for the wasps which pollinate a, *Ficus paraensis*; b, *F. popenoei*; and c, *F. trigonata*; in which 0.97, 0.36 and 0.07 of the broods respectively are founded by one mother, plotted with the observed mean sex ratios for one- and two-foundress broods. Deviations of the observed brood sex ratios from the predictions for one- (X) and two-foundress (Y) broods appropriate for each wasp species are indicated, as well as the observed shift in sex ratio between the one and two foundress broods (Z). Note that for broods with a given number of foundresses, the deviations from predicted are greatest in the species that rarely encounter that number of foundresses. Also, the observed shift in sex ratio from one to two foundress broods is greatest in the species in which neither situation is very rare.

one-foundress broods, $R_s = 0.747$, $P < 0.005$, for two-foundress broods; Fig. 2a and b).

Further, the data show that the fig wasp species which rarely experience broods with more than one foundress (such as those on *Ficus paraensis* and *F. pertusa*) or with only single foundresses (such as *F. trigonata*) show less adjustment between those two situations than wasp species which more commonly experience both situations (Spearman rank correlation coefficient, $R_s = 0.750$, $P < 0.005$, see Figs 1 and 3). Of all 13 species, only the wasps of *F. paraensis*, which have the lowest

Table 1 Characteristics of fig wasp broods sampled in the wild

Fig (<i>Ficus</i>) species	No. of fruit crops sampled	Proportion of fruits with single-foundress broods	$n^\dagger$	Arithmetic mean of no. of foundresses	Single-foundress broods		Two-foundress broods	
					Mean proportion of males	No. of broods sampled	Mean proportion of males	No. of broods sampled
<i>F. paraensis</i>	15	0.97	1.01	1.03	0.056	28	0.079	25
<i>F. pertusa</i>	8	0.95	1.03	1.05	0.105	59	0.143*	10
<i>F. bullenei</i>	6	0.86	1.08	1.18	0.065	11	0.135*	7
<i>F. obtusifolia</i>	17	0.83	1.10	1.21	0.052	45	0.118	17
<i>F. citrifolia</i>	21	0.80	1.11	1.24	0.056	29	0.146	50
<i>F. maxima</i>	10	0.70	1.32	1.48	0.106	17	0.170*	14
<i>F. yoponensis</i>	7	0.60	1.39	1.64	0.114	8	0.188*	8
<i>F. nymphitfolia</i>	12	0.57	1.44	2.32	0.054	27	0.195*	10
<i>F. dugandii</i>	7	0.54	1.75	2.35	0.106	8	0.187*	6
<i>F. popenoei</i>	16	0.36	1.82	2.33	0.099	8	0.200*	15
<i>F. insipida</i>	33	0.27	1.98	3.11	0.136	54	0.288	21
<i>F. near trigonata</i>	9	0.35	2.05	2.53	0.107	17	0.178	25
<i>F. trigonata</i>	10	0.07	3.46	4.68	0.203	14	0.250*	9

Each fruit crop represents 20 individual fruits sampled at random.

† Inverse of the estimated proportion of sib-mated females.

* Predicted optimal brood sex ratio is within the 95% confidence interval of the observed mean sex ratios.

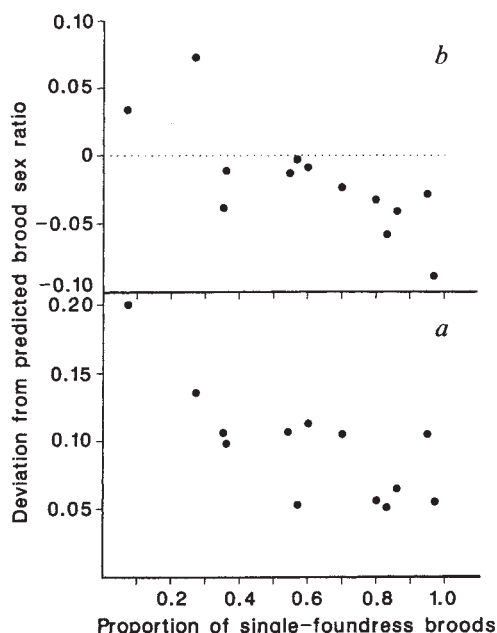


Fig. 2 *a*, Deviation from the predicted optimal brood sex ratio for broods with single foundresses plotted against the proportion of all broods with single foundresses in 13 species of fig pollinating wasps. The species in which more broods are founded by one mother show the least deviations from the predicted brood sex ratios for one mother. *b*, Deviation from the predicted optimal brood sex ratio for broods with two foundresses plotted against the proportion of all broods with single foundresses in 13 species of fig-pollinating wasps. As the proportion of multiple foundress broods increases, the deviations of observed brood sex ratios for two foundress broods from predicted optima become less female-biased until finally in the species in which single foundress broods are rarest, the deviations become male-biased.

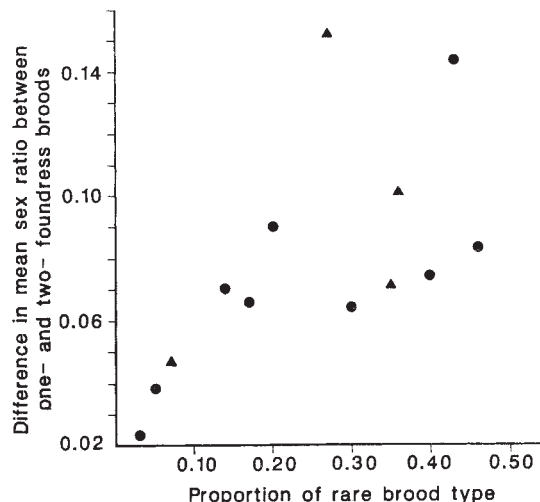


Fig. 3 The difference between one- and two-foundress brood sex ratios plotted against the proportion of the rarer brood type (single or multiple foundresses) in each of the 13 species. The species in which either single (▲) or multiple (●) foundress broods are rare show less sex ratio shift between one and two foundress broods than the species in which both of the two types of broods is more common.

proportion of the rare brood type (two-foundress broods, 0.03), show no significant difference between the sex ratios with one and two foundresses (t -Test, $t = 1.57$, d.f. = 37.6, $P = 0.125$, variances unequal). The predicted advantage of the optimum over the observed sex ratio in this two-foundress case is ~8% (ref. 7).

The data illustrate several points about sex ratio evolution in particular and adaptation in general. First, the most basic prediction of sex ratio theory, that sex ratios in subdivided populations are female-biased, is met. Further, the prediction that the brood sex ratios increase with increasing numbers of foundresses is met in the thirteen species of fig wasps. Also more inbred species show more female-biased sex ratios for any given number of foundresses (Spearman rank correlation coefficient, $R_s = 0.706$, $P < 0.01$, $n = 13$, for one-foundress broods and $R_s = 0.879$, $P < 0.001$, $n = 13$, for two-foundress broods). These results support the logical structure underlying the theory and the applicability of that theory to the natural history of the fig wasps. Second, the deviations from the predicted optima are not random. The wasps' responses fit theory best for those situations they encounter most frequently. Finally, although the cues and mechanisms are not understood, the wasps' ability to alter brood sex ratio in response to different LMC situations is most pronounced in the species that most frequently encounter variation between those situations. That is, adaptive plasticity in the wasps' sex ratio responses to different LMC situations is greater in more variable selective regimes. This pattern suggests a behavioural analogue to studies that document greater plasticity for physiological traits in populations that have historically experienced greater levels of environmental variation. Examples include different temperature tolerance and acclimatization abilities in

coastal versus desert plant populations¹⁷, or different osmoregulatory capabilities in different species or populations of crabs from fresh, brackish, or marine habitats¹⁸. An interesting difference is that in the cases of the plants and the crabs, individuals tend to experience a large portion of the range of possible environments over the course of their lifetime, whereas individual fig wasps only experience one LMC situation during their lives. Nonetheless, all these studies suggest that organisms rarely respond optimally across all environments that they may encounter. Instead, the specific adaptive responses are generally most appropriate for the specific, most frequently encountered situations and adaptive plasticity is most developed in the most variable selective regimes.

I thank T. P. Young, A. P. Smith, S. H. Orzack, and J. H. Werren for discussion; E. Jaen, R. Mihalik, and E. Rodriguez for field assistance; D. H. Feener, E. G. Leigh, H. F. Howe, R. D. Milkman, J. J. Bull, M. C. Singer, M. J. Ryan, P. J. Weldon, J. W. Gregg, J. K. Zimmerman and the Iowa Working Seminar for comments on the manuscript; R. B. Foster for identification of *Ficus* species, B. Minor for statistical advice; and D. S. Webster. This work was supported by the Smithsonian Institution and the University of Iowa.

Received 28 January; accepted 24 August 1987.

1. Charnov, E. L. in *The Theory of Sex Allocation* (Princeton University Press, 1982).
2. Leigh, E. G., Herre, E. A. & Fischer, E. A. *Experientia* **41**, 1265-1276 (1985).
3. Endler, J. A. in *Natural Selection in the Wild* (Princeton University Press, 1986).
4. Hamilton, W. D. *Science* **156**, 477-488 (1967).
5. Fisher, R. A. in *The Genetical Theory of Natural Selection* (Clarendon, Oxford, 1930).
6. Hamilton, W. D. in *Sexual Selection and Reproductive Competition in Insects* (eds Blum, M. S. & Blum, N. A.) 107-200 (Academic, New York, 1979).
7. Herre, E. A. *Science* **228**, 896-898 (1985).
8. Frank, S. A. *Evolution* **39**, 949-964 (1985).
9. Ramirez, B. W. *Science* **163**, 580-581 (1969).
10. Ramirez, B. W. *Evolution* **24**, 680-691 (1970).
11. Wiebes, J. T. A. *Rev. Ecol. Syst.* **10**, 1-12 (1979).
12. Taylor, P. D. & Bulmer, M. G. *J. theor. Biol.* **86**, 409-419 (1980).
13. Nunney, L. *Evolution* **39**, 349-361 (1985).
14. Stubblefield, J. W. thesis, Harvard Univ. (1980).
15. Kjellberg, F. thesis, Inst. National Agronomique, Paris (1983).
16. Joshi, N. V. & Gadagkar, R. *J. Genet.* **64**, 41-58 (1985).
17. Berry, J. A. & Raison, R. K. in *Encyclopedia of Plant Physiology*, **12a** (eds Lange, O. L., Nobel, P. S., Osmond, C. B. & Ziegler, H.) (Springer, New York, 1981).
18. Mantel, L. H. & Farmer, L. L. in *The Biology of the Crustacea* Vol. 5 (ed. Bliss, D. E.) 53-161 (Academic, New York, 1983).

Macrophage-induced angiogenesis is mediated by tumour necrosis factor- α

S. Joseph Leibovich*, Peter J. Polverini†, H. Michael Shepard‡, David M. Wiseman*, Vera Shively* & Nureddin Nuseir*

Departments of * Oral Biology and † Pathology, Northwestern University Dental School, 303 E. Chicago Avenue, Chicago, Illinois 60611, USA

‡ Department of Pharmacological Sciences, Genentech Inc., South San Francisco, California 94080, USA

Macrophages are important in the induction of new blood vessel growth during wound repair, inflammation and tumour growth¹⁻⁴. We show here that tumour necrosis factor- α (TNF- α), a secretory product of activated macrophages that is believed to mediate tumour cytotoxicity⁵⁻⁹, is a potent inducer of new blood vessel growth (angiogenesis). *In vivo*, TNF- α induces capillary blood vessel formation in the rat cornea and the developing chick chorioallantoic membrane at very low doses. *In vitro*, TNF- α stimulates chemotaxis of bovine adrenal capillary endothelial cells and induces cultures of these cells grown on type-1 collagen gels to form capillary-tube-like structures. The angiogenic activity produced by activated murine peritoneal macrophages is completely neutralized by a polyclonal antibody to TNF- α , suggesting immunological features are common to TNF- α and the protein responsible for macrophage-derived angiogenic activity. In inflammation and wound repair, TNF- α could augment repair by stimulating new blood vessel growth; in tumours, TNF- α might both stimulate tumour development by promoting vessel growth and participate in tumour destruction by direct cytotoxicity¹⁰⁻¹².

Figure 1 shows the strong and sustained growth of new capillary blood vessels from the corneal limbus toward a Hydron implant containing 3.5 ng TNF- α implanted in the cornea of a rat. Below 350 ng corneas showed no clouding or oedema, indicating that inflammation was not a significant component of the angiogenic reaction. This was confirmed by histological examination, which also showed an absence of infiltrating leukocytes. Pellets containing more than 350 ng induced mild, transient oedema, but no haemorrhagic exudate was observed. Figure 2 shows the strong neovascular response induced by TNF- α on the developing chick chorioallantoic membrane (CAM). TNF- α (1 ng) induced a large increase in the density of the microvascular bed in the region of the pellets, with apparent regression of larger vessels, and 5–50 ng induced a

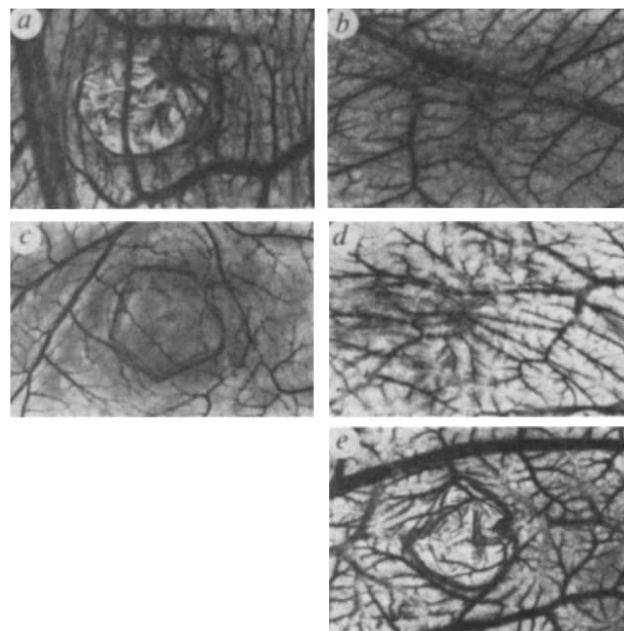


Fig. 2 Neovascular responses induced on chick chorioallantoic membranes of shell-free egg cultures. *a*, 1 ng TNF- α ; *b*, 5 ng TNF- α ; *c*, negative response of medium control; *d*, positive response of macrophage-conditioned medium; *e*, negative response of macrophage-conditioned medium treated with antibody to TNF- α ($\times 11$).

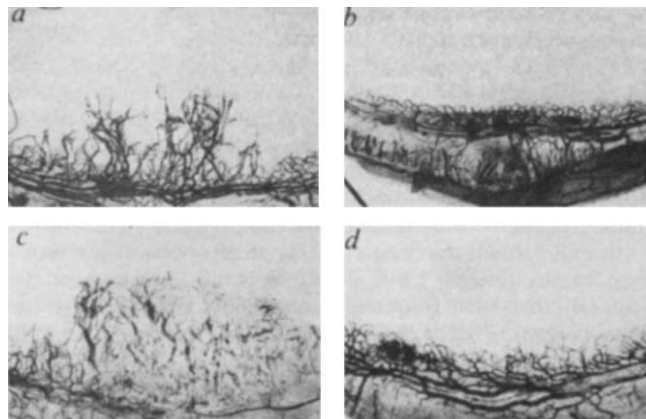
Methods. Fertilized chick eggs were cracked on the third day of gestation into plastic-wrap cradles²⁹. The shell-free embryos developed normally for 7 days at 37 °C in a humidified incubator. One vol. test substance was mixed with 9 vols 0.5% (w/v) methyl cellulose in water and 10 μ l applied to the square cut ends of 2 mm Teflon rods and dried. These pellets were positioned on the developing chorioallantoic membrane of the shell-free egg cultures. Growth of new blood vessels was monitored daily for 3 d using a Wild M5A stereomicroscope (Heerbrug). Antibody treatment of macrophage-conditioned medium was as described in Table 1 legend.

characteristic spokewheel pattern of new vessel growth. At higher concentrations, inflammatory responses were evident.

In vitro, in a modified Boyden microwell assay system, TNF- α potently induced chemotaxis of bovine adrenal capillary endothelial cells (BCEs) across gelatin-coated polycarbonate filters (Table 1). Peak stimulation of chemotactic activity was observed at 5–50 ng ml⁻¹ TNF- α , but activity was seen down to 0.5 ng ml⁻¹. The maximal migratory response was the same or

Fig. 1 Neovascular responses induced in corneas of rats 7 d after implantation of Hydron pellets. *a*, Positive neovascular response induced by 3.5 ng (0.2 pmoles) TNF- α ; *b*, negative response of control implant; *c*, positive response induced by thioglycollate-induced murine peritoneal macrophage-conditioned medium; *d*, negative response of macrophage-conditioned medium treated with rabbit antibody to TNF- α ($\times 15$).

Methods. Murine recombinant TNF- α (ref. 9) ($\sim 2.9 \times 10^7$ U mg⁻¹) and polyclonal rabbit antibody to murine rTNF- α (1,325 neutralizing units μ l⁻¹) were prepared at Genentech. The activity of TNF- α is based on its cytotoxicity toward murine L-M fibroblasts in the presence of actinomycin D. One unit of TNF- α is defined as the reciprocal of the test dilution resulting in 50% cytotoxicity. Test samples at the appropriate dilution were mixed with equal volumes of Hydron (12% w/v in 95% ethanol) and 10 μ l placed on the square cut ends of 2 mm diameter teflon rods (Berghoff) and allowed to dry under reduced pressure. These pellets were implanted in the corneas of F344 rats, at ~ 1.5 mm from the corneal limbus^{2,4}. Corneas were monitored daily for ingrowth of new microvessels from the limbal vasculature toward the implants. After 7 d, animals were perfused intra-arterially with colloidal carbon (Pellikan, FRG), and the corneas fixed and excised to obtain a permanent record of the vascular pattern of growth. Antibody treatment of macrophage-conditioned medium is described in the legend to Table 1.



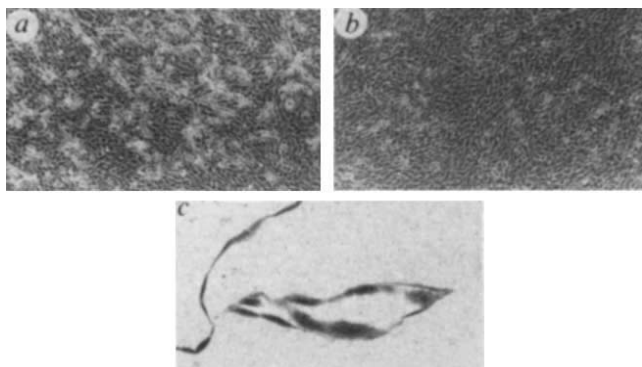


Fig. 3 Capillary-tube-like structures formed from BCEs cultured on collagen gels. *a*, Capillary-tube-like structures induced in 24 h by 5 ng ml⁻¹ TNF- α ($\times 53$); *b*, control cultures in the absence of TNF- α ; *c*, light micrograph of a transverse section of capillary tubes shown in *a* ($\times 215$).

Methods. A modified assay³⁰ was used to test samples for their ability to induce BCE monolayers to organize into capillary-tube-like structures that invade collagen gels. Collagen gels were prepared in 12-well culture plates (Costar) by combining 8 vols bovine dermal collagen solution (Vitrogen) (3.6 mg ml⁻¹ in 0.012 M HCl) with 1 vol. 10 $\times$ RPMI1640 and 1 vol. 7.5% NaHCO₃, diluted 1:10 with 0.142 M NaOH. Buffered collagen (0.5 ml) was allowed to gel overnight at 37 °C in the wells under 95% O₂/5% CO₂. BCEs were isolated and cloned from bovine adrenal cortex³¹ and cultured in DMEM-20% FCS. 1 ml BCEs (2 $\times$ 10⁶ ml⁻¹) at passage 11 or 12 was seeded into each well. Plates were incubated until cultures reached confluence (2–3 days), then washed and treated with TNF- α or control media. After 24 h the cultures were washed with PBS, fixed with 2.5% paraformaldehyde-1% glutaraldehyde in cacodylate buffer and examined in the phase contrast microscope; selected cultures were embedded in glycol methacrylate and processed for light microscopy. Bovine bFGF (R & D Systems) was a positive control.

greater than that induced by 10 ng ml⁻¹ basic fibroblast growth factor (bFGF). Also, confluent cultures of BCEs showed distinct changes in morphology when cultured on collagen gels in the presence of TNF- α (Fig. 3). TNF- α at 50 ng ml⁻¹ down to 3.5 ng ml⁻¹ induced the formation of branching capillary-tube-like structures that grew from the endothelial monolayer into the collagen gels within 24–48 hours (Fig. 3). At higher concentrations of TNF- α (above 350 ng ml⁻¹), some rounding of cells, followed by shedding from the gels, was observed after 24–48 hours incubation but was not seen at lower concentrations.

To determine if the angiogenic activity produced by activated macrophages in culture was related to TNF- α , a polyclonal antibody to murine TNF- α was used. This antibody completely neutralized the angiogenic activity in conditioned medium of thioglycollate-induced peritoneal macrophages, indicating that the macrophage-derived angiogenic agent is either identical or immunologically closely related to TNF- α . This neutralization of activity was demonstrated in the rat cornea, the chick CAM, and the BCE chemotactic and capillary tube formation assays.

These results indicate that TNF- α is a mediator of angiogenesis, with activity at concentrations lower than 3.5 ng (~0.2 pmoles) per implant in both the rat cornea and the chick CAM. This compares with published reports for induction of corneal neovascularization for acidic and basic fibroblast growth factor (FGF) of 0.33–0.54 pmoles (refs 13–15), and for angiogenin of 3.5 pmoles¹⁶. Epidermal growth factor (EGF) (2 nmoles) and transforming growth factor- α (TGF- α) (50–170 pmoles) were shown to be angiogenic in the hamster cheek pouch¹⁷. Angiogenic activity of these factors in the cornea has not been reported, making direct comparison difficult. Nevertheless, TNF- α appears to be angiogenic at concentrations comparable to, or considerably lower than, those reported for FGF, angiogenin, EGF and TGF- α .

Table 1 Chemotaxis of bovine adrenal capillary endothelial cells (BCEs)

Concentration of test sample	No. BCEs ten high power fields ($\pm$ s.e.m.)
TNF- α :	
500 ng ml ⁻¹	48 $\pm$ 5
50 ng ml ⁻¹	79 $\pm$ 4.6
5 ng ml ⁻¹	51 $\pm$ 3.5
0.5 ng ml ⁻¹	31 $\pm$ 2
50 ng ml ⁻¹ + rabbit anti-TNF- α	24 $\pm$ 3
50 ng ml ⁻¹ + non-immune rabbit serum	76 $\pm$ 4
Control medium (RPMI1640-1% FCS)	22 $\pm$ 4
bFGF: 10 ng ml ⁻¹	72 $\pm$ 13
Macrophage-conditioned medium (MCM)	77 $\pm$ 7
MCM + rabbit anti-TNF- α	25 $\pm$ 4.5

Chemotaxis of BCEs was performed in 25 μ l modified blind well chambers (Neuroprobe). Polycarbonate filters (8 μ m pore diameter) were soaked in 0.5 M acetic acid, washed, incubated for 24 h in 100 μ g calf skin gelatin per ml (Type III, Sigma) and air-dried. TNF- α in RPMI1640-1% FCS in the lower wells was incubated with 50 μ l BCE suspension (2.5 $\times$ 10⁶ cells ml⁻¹ RPMI1640-1% FCS) in the top wells of the chamber for 4 h at 37 °C. Filters were fixed in methanol and stained with Diff-Quik (American Scientific Products). The number of cells migrating to the lower surface of the filter was counted in ten high power ($\times 400$) fields per well. Results are expressed as the mean $\pm$ standard error of the mean (SEM) for triplicate determinations. bFGF (10 ng ml⁻¹) was a positive chemotactic control. Macrophage-conditioned medium (MCM) was prepared using thioglycollate-induced peritoneal macrophages from C57/B1 mice (7–10 weeks)², and concentrated tenfold. For antibody treatment of TNF- α or MCM, 1–5 μ l rabbit anti-murine TNF- α polyclonal antibody was added to the incubation at 22 °C, continued for 4 h, then for 2 h with 10 ml immunobeads (BioRad, which consisted of goat anti-rabbit immunoglobulin, heavy- and light-chain specific, bound to polyacrylamide beads. The immunobeads were spun off, the supernatants concentrated tenfold, dialysed against DMEM, and then assayed for angiogenic, chemotactic and capillary-tube-forming activity. Control incubations used normal, non-immune rabbit serum. The rabbit anti-TNF- α antibody is highly specific, does not cross react with FGF, and did not affect the activity of bFGF in any of the systems tested.

TNF- α is a major secretory product of activated macrophages^{5–9}, and has been implicated as the macrophage product inducing tumour cell cytotoxicity. TNF- α is identical to cachectin, a macrophage product thought to be responsible for the wasting response associated with some bacterial and parasitic infections¹⁸. TNF- α has several effects on human endothelial cells, inducing their expression of class I major histocompatibility antigens, procoagulant activity, and interleukin-1, as well as increasing their adhesiveness for polymorphonuclear leukocytes^{5,8,19}. Purified rabbit TNF- α inhibits growth and induces distinct morphological changes in cultured microvascular endothelial cells²⁰. It is rapidly becoming apparent that TNF- α is a pleiotropic mediator with wide-ranging biological effects, not only of tumour cytotoxicity, but also of inflammation and immune reactivity^{5,8,19}. Expression of TNF- α by monocytes requires activation of the cells, for example with γ -interferon and endotoxin²¹. We showed earlier that expression of angiogenic activity by monocytes also requires activation²² and, more recently²³, that transforming growth factor- β (TGF- β), a peptide product of numerous transformed and normal cells, is strongly chemotactic for blood monocytes and can induce their expression of angiogenic activity. Platelet α -granules are a rich source of TGF- β (ref. 24), which when released locally at sites of injury could stimulate recruitment of monocytes and their subsequent activation to produce angiogenic substances, such as TNF- α .

FGF is a potent angiogenic agent present in almost all normal adult tissues²⁵ and found in endothelial cells²⁶. Endothelial

proliferation in these tissues in the absence of inflammation, however, is generally exceedingly low, with turnover time measured in years²⁷. FGF is also present in²⁸, but is not secreted by, macrophages (S.J.L. & P.J.P., unpublished observations). The mechanism of action of TNF- α in inducing angiogenesis is not yet understood. It is active at concentrations as low as those associated with FGF activity. As a secretory product of activated macrophages, TNF- α is a candidate for an inducer of angiogenesis in inflammation, wound repair and tumour growth.

This work was supported in part by funds from Northwestern University Dental and Medical Schools. D.M.W. was the recipient of a Wellcome Travel Grant.

Received 3 June; accepted 3 September 1987.

1. Leibovich, S. J. & Ross, R. *Am. J. Path.* **78**, 71–100 (1975).
2. Polverini, P. J., Cotran, R. S., Gimbrome, M. A. Jr & Unanue, E. R. *Nature* **269**, 804–806 (1977).
3. Clark, R. A. *et al. Surg. Forum* **27**, 16–22 (1976).
4. Polverini, P. J. & Leibovich, S. J. *Lab. Invest.* **51**, 635–642 (1985).
5. Nathan, C. F. *J. clin. Invest.* **79**, 319–326 (1987).
6. Carswell, E. A. *et al. Proc. natn. Acad. Sci. U.S.A.* **72**, 3666–3670 (1975).
7. Ruff, M. & Gifford, G. E. in *Lymphokine Reports*, 2 (ed. Pick, E.) 235–272 (Academic, New York, 1985).
8. Beutler, B. & Cerami, A. *Nature* **320**, 584–588 (1986).

9. Urban, J. L., Shepard, H. M., Rothstein, J. L., Sugarman, B. J. & Schreiber, H. *Proc. natn. Acad. Sci. U.S.A.* **83**, 5233–5237 (1986).
10. Sugarman, B. J. *et al. Science* **230**, 943–945 (1985).
11. Banda, M. J., Knighton, D. R., Hunt, T. K. & Werb, Z. *Proc. natn. Acad. Sci. U.S.A.* **79**, 7773–7777 (1982).
12. Koch, A. E., Polverini, P. J. & Leibovich, S. J. *Arthritis Rheum.* **29**, 471–479 (1986).
13. Shing, Y. *et al. J. cell. Biochem.* **29**, 275–287 (1985).
14. Thomas, K. A., Rios-Candelore, M., Gimenez-Gallego, G., Rodkey, J. & Fitzpatrick, S. *Proc. natn. Acad. Sci. U.S.A.* **82**, 6409–6413 (1985).
15. Esch, F. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 6507–6511 (1985).
16. Fett, J. W. *et al. Biochemistry* **24**, 5480–5486 (1985).
17. Schreiber, A. B., Winkler, M. E. & Derynck, R. *Science* **232**, 1250–1252 (1986).
18. Beutler, B. *et al. Nature* **316**, 552–554 (1985).
19. Le, J. & Vilcek, J. *Lab. Invest.* **56**, 234–248 (1987).
20. Sato, N. *et al. Proc. Japan Acad.* **61**, 471–474 (1985).
21. Beutler, B., Tkachenko, V., Milsark, I., Krochin, N. & Cerami, A. *J. exp. Med.* **164**, 1791–1796 (1986).
22. Koch, A. E., Polverini, P. J. & Leibovich, S. J. *J. Leukocyte Biol.* **39**, 233–238 (1986).
23. Wiseman, D., Polverini, P. J., Kamp, D. & Leibovich, S. J. *J. Cell Biol.* (in the press).
24. Assoian, R. K. & Sporn, M. B. *J. cell Biol.* **102**, 1217–1223 (1986).
25. Klagsburn, M., Sasse, J., Sullivan, R. & Smith, J. A. *Proc. natn. Acad. Sci. U.S.A.* **83**, 2448–2452 (1986).
26. Schweigerer, L. *et al. Nature* **325**, 257–259 (1987).
27. Denekamp, J. in *Progress Microcirculation* Vol. 42 (eds Hammersen, F. & Hudlicka, O.) 28–38 (Karger, Basel, 1984).
28. Baird, A., Mormede, P. & Bohlen, P. *Biochem. biophys. Res. Commun.* **126**, 358–364 (1985).
29. Montesano, R. & Orci, L. *Cell* **42**, 469–477 (1985).
30. Leibovich, S. J. & Polverini, P. J. *Br. J. Rheumatol.* **24**, 197–202 (1985).
31. Folkman, J., Haudenschild, C. C. & Zetter, B. R. *Proc. natn. Acad. Sci. U.S.A.* **76**, 5217–5221 (1979).

Dependence on pH of polarized sorting of secreted proteins

Michael J. Caplan, Jennifer L. Stow, Anna P. Newman, Joseph Madri*, H. Clarke Anderson†, Marilyn G. Farquhar, George E. Palade & James D. Jamieson

Department of Cell Biology and * Department of Pathology, Yale University School of Medicine, 333 Cedar Street, New Haven, Connecticut 06510, USA

† Kansas University Medical Center, Rainbow Boulevard at 39th Kansas City, Kansas 66103, USA

The plasma membranes of epithelial cells are divided into apical and basolateral domains. These two surfaces are characterized by markedly different protein compositions, reflecting the ability of the cell to target newly synthesized membrane proteins to specific regions of the cell surface^{1,2}. This targeting capability is also apparent in the polarized release of secretory products. Recent studies using canine renal tubule (MDCK) cells have suggested that distinct sets of secretory proteins are released from their apical and basolateral poles^{3–6}. We report experiments designed to examine secretory protein sorting by MDCK cells. We have shown that secretion of basement membrane components (laminin and heparan sulphate proteoglycan (HSPG)) takes place from the basolateral cell surface and that this polarized release results from active sorting. The sorting process which mediates this polarized secretion requires an acidic intracellular compartment. MDCK cells treated with NH₄Cl to raise the pH of their intracellular compartments^{7–9}, secrete laminin and HSPG by a default pathway which leads to their release in roughly equal quantities into the medium of both the apical and basolateral compartments.

Before the prerequisites for polarized secretion of basement membrane components could be defined, it was necessary to determine the fate of secreted polypeptides which lack sorting signals and are unable to interact with the sorting apparatus. The route pursued by such proteins defines the default pathway of the cell—that is, the pathway available to secretory proteins which are not substrates for active posttranslational sorting. In principle, there were three distinct possibilities: unsorted proteins could be secreted solely apically, solely basolaterally, or both apically and basolaterally. In defining the default pathway, the actively sorted pathways are implicitly delineated. For example, if default secretion were to occur apically, then baso-

lateral release of polypeptides would require active sorting; if the default pathway were to carry unsorted proteins to both the apical and basolateral surfaces, then exclusive secretion of a polypeptide at either surface would implicate two different sorting mechanisms.

To examine the secretion of proteins that are not sorted, polypeptides which lack sorting signals or are prevented from interacting with the sorting mechanism must be identified or generated. Others^{3,4} have found that MDCK cells transfected with the genes for foreign secretory proteins release these polypeptides both apically and basolaterally. Because these exogenous proteins may be incapable of interacting with sorting apparatus of the MDCK cell, these authors postulate that release by the default pathway results in the observed non-polarized secretion. We chose to study the default pathway by preventing the sorting of lysosomal enzymes. In cells which have been treated with NH₄Cl, a weak base that raises the pH of intracellular acidic compartments^{7–9}, newly synthesized lysosomal enzymes cannot interact with their appropriate sorting receptors during passage through the Golgi complex, and are thus released constitutively into the extracellular medium^{10–13}. Examining the polarity of secretion of newly synthesized lysosomal enzymes from NH₄Cl-treated cells should, therefore, provide an alternative approach to identifying the pathway for default secretion.

To study the polarity of protein secretion, MDCK cells were grown to confluence on polycarbonate Nuclepore filters mounted in mini-Marbrook chambers¹⁴. Under these conditions they form monolayers with intercellular tight junctions¹ that can prevent inulin and ouabain (of relative molecular mass, M_r , 4,000 and 728, respectively) from crossing the monolayer¹⁵. The plasma membrane polarity of these monolayers has been established by previous studies (ref. 15; and see ref. 1 for review). Monolayer polarity and impermeability are not affected by the addition of 10 mM NH₄Cl (ref. 15).

Filter-grown MDCK cells were pulse labelled with ³⁵S-methionine which was added to the basolateral medium compartment as described in Fig. 1. The apical and basolateral media were immunoprecipitated with a polyclonal antibody directed against porcine cathepsin D (ref. 16). MDCK cells treated with 10 mM NH₄Cl throughout the labelling period secrete newly synthesized cathepsin D equally into both the apical and basolateral media (Fig. 1). As expected, omission of NH₄Cl results in release of little or no newly synthesized enzyme¹³. Similarly, washout of NH₄Cl from pretreated monolayers prevents discharge of radiolabelled cathepsin D into either compartment (data not shown). These findings support the proposition that unsorted secreted proteins, exemplified by

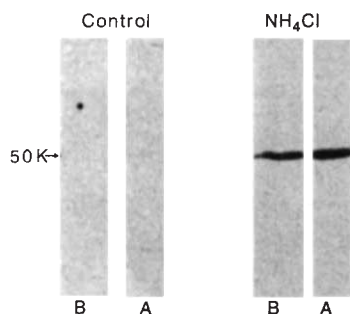


Fig. 1 Secretion of cathepsin D from MDCK monolayers. MDCK cells, grown to confluence on Nuclepore filters, were labelled for 3 h with ^{35}S -methionine in the presence or absence of 10 mM NH_4Cl . Following labelling, apical (lanes A) and basolateral (lanes B) media were immunoprecipitated with a polyclonal antibody against porcine cathepsin D (ref. 16). In controls incubated without NH_4Cl (left-hand two lanes), little or no newly synthesized cathepsin D can be detected in fluorographs of the immunoprecipitates obtained from either the apical or basolateral medium. Addition of 10 mM NH_4Cl results in secretion of the prelysosomal 50K form of cathepsin D into both compartments (right-hand lanes A and B) in roughly equal proportions.

Methods. MDCK cells (strain II, low resistance¹⁵) were passaged and maintained on Nuclepore filters in Eagle's minimal essential medium with Earle's salts and supplemented with 5% fetal calf serum (EMEM) as previously described¹⁵. Before radiolabelling, filter chambers bearing monolayers were washed twice in methionine-free EMEM (20 mM HEPES, pH 7.4) and were incubated for 3 h at 37°C with their apical surfaces bathed in 1.5 ml of methionine-free EMEM (20 mM HEPES) and their basolateral surfaces in contact with 1.5 ml of methionine-free EMEM (20 mM HEPES) supplemented with 0.5 mCi of ^{35}S -methionine (>1,000 Ci mmol⁻¹, Amersham). Immunoprecipitation was performed on 0.5 ml of apical or basolateral medium as previously described¹⁵. Immunoprecipitates were analysed by SDS polyacrylamide gel electrophoresis (PAGE) on 7.5% gels²⁶ which were processed for fluorography using Autofluor (National Diagnostics). All experiments reported are representative of at least three trials.

unsorted cathepsin D, leave the cell both apically and basolaterally. This description of the default pathway is consistent with that derived from the transfection experiments described above^{3,4}.

The data presented in Fig. 1 indicate that newly synthesized proteins which are secreted preferentially at one or the other surface of the MDCK monolayer require active sorting to attain their polarized release. We hypothesized that likely candidates for polarized secretion might be found among the proteins making up the MDCK cell basement membrane. Epithelial cells deposit a basement membrane which underlies their basal plasmalemma^{17,18}. Although the constituents of basement membranes are assembled only at the basal pole of the cell, it has yet to be determined whether their secretion is similarly polarized. We therefore examined the synthesis and secretion of two basement membrane polypeptides: laminin¹⁹ and HSPG (ref. 20).

Filter-grown MDCK monolayers were again radiolabelled with ^{35}S -methionine as in Fig. 1. After labelling, sonicates of cells were prepared and, along with the apical and basolateral media, were immunoprecipitated with polyclonal antibodies against laminin²¹ or rat kidney basement membrane HSPG (ref. 22). Radiolabelled proteins corresponding to immature forms of laminin²³ were immunoprecipitated from the cell homogenates (Fig. 2). The fully processed form of laminin was present only in the basolateral medium; no laminin polypeptides were detected in the apical medium. Similarly, newly synthesized HSPG was released predominantly at the basolateral surface (Fig. 4, 'control'). It seems that laminin and HSPG are not released randomly from polarized MDCK cells but instead are preferentially directed to a basolateral secretory pathway.

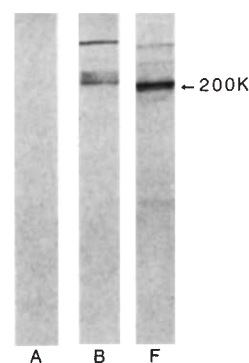


Fig. 2 Polarized secretion of laminin from MDCK monolayers. MDCK cells grown to confluence on Nuclepore filters were labelled with ^{35}S -methionine for 3 h. After labelling, apical and basolateral media were collected, filters were rinsed twice in NET buffer (150 mM NaCl, 2 mM EDTA, 10 mM Tris HCl pH 7.5), and cellular contents were extracted by briefly sonicating the filters (30 s, Branson microtip, 40% output) in 1.5 ml NET buffer supplemented with 1% Nonidet P-40. Immunoprecipitation with a polyclonal antibody against EHS sarcoma laminin was performed on 0.5 ml of the apical (lane A) or basolateral (lane B) medium and on 0.5 ml of cell extracts (lane F). Fluorographic analysis of immunoprecipitates separated by SDS-PAGE reveals that little or no newly synthesized laminin can be detected in apical media. In contrast, the complex of bands characteristic of fully processed laminin²³ was readily detectable in media from the basolateral compartment. Newly synthesized laminin was also present in the cell extract. The pattern of bands depicted in lane F corresponds to intracellular processing intermediates of the two laminin subunits²³. The data demonstrate that this basement membrane component is released solely into the basolateral compartment. Furthermore, under our growth conditions the ratio of basolateral to apical surface area is ~2:1 (data not shown). Thus, together with the results of the cathepsin D experiment (Fig. 1), this finding suggests that an active sorting mechanism functions to target newly synthesized laminin to the basolateral domain of the cell.

The basolateral secretion of laminin and HSPG must require active sorting. These two basement membrane components presumably interact with some targeting mechanism which mediates their diversion from the default pathway and results in their polar discharge. A possible mechanism is suggested by work with the non-polar AtT-20 pituitary cell line in which some secretory proteins pursue a course leading to constitutive release whereas others follow a regulated pathway resulting in their packaging into storage granules whose discharge is stimulated by secretagogues²⁴. Sorting to the regulated pathway seems to require an acidic intracellular compartment. AtT-20 cells treated with NH_4Cl are unable to target proteins to secretory granules and thus release their newly synthesized secretory proteins constitutively. We wished to test the possibility that an acidic intracellular compartment is also necessary for the sorting of basolateral secretory proteins in polarized MDCK cells.

Confluent, filter-grown MDCK cell monolayers were incubated with ^{35}S -methionine in the presence or absence of 10 mM NH_4Cl . Apical and basolateral media were immunoprecipitated with antibodies against laminin and HSPG. When NH_4Cl is omitted (Fig. 3, 'control'), newly synthesized laminin is detected only in the basolateral medium. MDCK cells treated with 10 mM NH_4Cl during radiolabelling, however, secrete newly synthesized laminin equally into the apical and basolateral media. NH_4Cl has a similar effect on HSPG secretion (Fig. 4). Substantial quantities of newly synthesized HSPG are detected in the apical as well as the basolateral media. The effect of NH_4Cl on the sorting of laminin is reversible, because washing this drug out of cells before radiolabelling results in a secretory pattern indistinguishable from that of untreated cells (Fig. 3, 'washout').

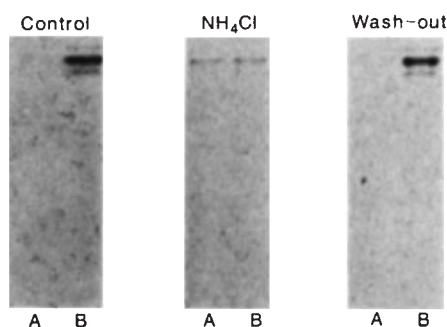


Fig. 3 Weak bases reversibly block the proper sorting of newly synthesized laminin. MDCK cells grown to confluence on Nucleopore filters were labelled for 3 h with ^{35}S -methionine in the presence or absence of 10 mM NH_4Cl . Apical (lanes A) and basolateral (lanes B) media were then immunoprecipitated with antibody against laminin followed by SDS-PAGE and fluorography of the immunoprecipitates. In the absence of NH_4Cl ('control', left-hand two lanes), newly-synthesized laminin could only be immunoprecipitated from the basolateral medium compartment, in keeping with the results depicted in Fig. 2. The same complex of three bands that is depicted in the B lanes in Fig. 2 is detected, although with slightly different mobilities in this electrophoretogram. When labelling was carried out in the presence of 10 mM NH_4Cl (' NH_4Cl ' centre two lanes), laminin could be demonstrated in both the apical and basolateral media in roughly equal proportions. MDCK monolayers which were incubated with 10 mM NH_4Cl for 3 h and subsequently metabolically labelled for 3 h with ^{35}S -methionine in the absence of NH_4Cl ('wash-out', right-hand two lanes) secreted newly synthesized laminin polypeptides basolaterally. Thus, NH_4Cl blocks the normal sorting of newly synthesized laminin polypeptides, resulting in their secretion by the default pathway into both the apical and basolateral medium. This effect is fully reversible, because washing the NH_4Cl out of the cells before radiolabelling restores polarized laminin secretion.

NH_4Cl does not seem to affect the sorting of apically secreted polypeptides. As reported earlier, a polypeptide with M_r 81,000 (81K) is the major labelled species released into the apical medium by MDCK cells incubated with ^{35}S -methionine⁴. In the absence of protease inhibitors, proteolytic fragments of this polypeptide, with M_r 30, 35 and 40K, are detected^{3,4}. We find this complex of three bands predominantly in the apical medium collected from filter-grown MDCK cells labelled with ^{35}S -methionine in the presence or absence of 10 mM NH_4Cl (Fig. 5). Note that Gottlieb *et al.*⁴ find that the addition of chloroquine markedly retards secretion of the 81K polypeptide. We do not observe this effect in experiments using NH_4Cl .

It seems, therefore, that the sorting process that targets newly synthesized proteins for basolateral secretion involves an intracellular acidic compartment. Raising the pH of such a compartment with NH_4Cl disrupts the sorting mechanism and results in the release of laminin and HSPG by a default pathway (the pathway pursued by unsorted proteins) into both the apical and basolateral medium. In interpreting this data, note that NH_4Cl could induce other effects that might also account for our observations. For example, it might exert its effect not at the level of sorting but by randomizing intracellular vesicular traffic. We believe this latter possibility to be unlikely, because we have previously shown that 10 mM NH_4Cl does not disrupt vectorial delivery of newly synthesized $(\text{Na}^+ + \text{K}^+)\text{ATPase}$ to the MDCK cell basolateral plasmalemma¹⁵. Furthermore, the data presented in Fig. 5 suggest that NH_4Cl does not affect the sorting of proteins normally released into the apical medium. These results suggest that NH_4Cl perturbs the sorting of basolateral secretory proteins by an effect on the pH of some intracellular compartment.

The nature of this acidic compartment and its role in sorting remain unknown. It is tempting to postulate that the sorting of basolateral secretory proteins (that is basement membrane components) occurs by mechanisms similar to those that mediate

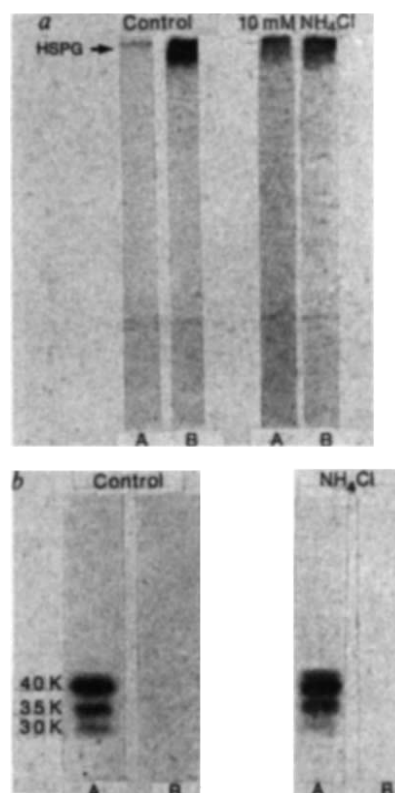


Fig. 4 *a*, HSPG is targeted primarily to the basolateral medium compartment by an NH_4Cl -sensitive pathway. MDCK cells grown to confluence on Nucleopore filters were labelled for 3 h with ^{35}S -methionine. Apical (lanes A) and basolateral (lanes B) media were immunoprecipitated with a polyclonal antibody against the core polypeptide of rat basement membrane HSPG (ref. 22). Immunoprecipitates were analysed by SDS-PAGE and fluorography. In the absence of NH_4Cl ('control' lanes), newly synthesized HSPG is secreted predominantly into the basolateral medium. Densitometric scanning of the control lanes in Fig. 4 reveals that >90% of the radiolabelled HSPG secreted by the MDCK monolayer was discharged into the basolateral medium. Inclusion of 10 mM NH_4Cl in the labelling medium results in substantial quantities of HSPG being secreted into the apical medium compartment. *b*, Sorting of apically secreted proteins is not affected by NH_4Cl . MDCK cells grown to confluence on Nucleopore filters were labelled for 3 h with ^{35}S -methionine. Aliquots of apical (lanes A) or basolateral (lanes B) medium were analysed by SDS-PAGE followed by fluorography. In the absence of NH_4Cl ('control' lanes at left), apical media contained a complex of three bands (M_r weights 30, 35 and 40K), as has been previously reported by others^{3,4}. The presence of 10 mM NH_4Cl (' NH_4Cl ' lanes at right) during the labelling incubation does not affect the predominantly apical release of these polypeptides. These results are consistent with the possibility that the targeting of proteins for apical secretion differs from basolateral sorting in its insensitivity to weak bases.

diversion of newly synthesized lysosomal enzymes to lysosomes¹⁰⁻¹³. Such a model requires that a membrane associated receptor binds these secretory proteins in a pH-sensitive manner and subsequently targets them to their final destination. A low-pH compartment might be required to terminate (or initiate) the sorting function of the receptor. We note that of the membrane proteins studied to date, which include influenza HA protein, destined for the apical domain²⁵, and the $(\text{Na}^+ + \text{K}^+)\text{ATPase}$, destined for the basolateral domain¹⁵, none is missorted by MDCK cells in the presence of NH_4Cl . Thus, the sorting mechanisms which operate for polarized delivery of membrane proteins are at least in part distinct from those which govern secretory protein sorting. Verification of this model of secretory protein sorting awaits the identification and characterization of the postulated sorting receptor.

We thank Drs T. Takahashi and J. Tang for the generous gift of antibody against cathepsin D, Dr I. Mellman for comments on the manuscript, G. Ouyang for technical assistance, and E. Lewis for help in preparing the manuscript and A. Curley-Whitehouse for help with the figures. This work was partially supported by NIH grants to J.D.J. and M.J.C.

Received 6 August; accepted 24 August 1987.

1. Simons, K. & Fuller, S. D. *A. Rev. Cell Biol.* **1**, 243-288 (1985).
2. Matlin, K. *J. Cell Biol.* **103**, 2565-2568 (1986).
3. Kondor-Koch, C., Bravo, R., Fuller, S. D., Cutler, D. & Garoff, H. *Cell* **43**, 297-306 (1985).
4. Gottlieb, T. A. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 2100-2104 (1986).
5. Gonzalez, A. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 3738-3742 (1987).
6. Roth, M. G., Gundersen, D., Patil, N. & Rodriguez-Boulant, E. *J. Cell Biol.* **104**, 769-782 (1987).
7. Maxfield, F. R. *J. Cell Biol.* **95**, 676-681 (1982).
8. Dean, R. T., Jessup, W. & Roberts, C. R. *Biochem. J.* **218**, 27-40 (1984).

9. Poole, B. & Ohkuma, S. *J. Cell Biol.* **90**, 665-669 (1981).
10. Brown, W. J., Constantinescu, E. & Farquhar, M. G. *J. Cell Biol.* **99**, 320-326 (1984).
11. Creek, K. E. & Sly, W. S. In *Lysosomes in Biology and Pathology* (eds Dingle, J. T., Dean, R. T. & Sly, W. S.) 63-82 (Elsevier, New York, 1984).
12. Brown, W. J., Goodhouse, J. & Farquhar, M. G. *J. Cell Biol.* **103**, 1235-1247 (1986).
13. Gonzalez-Noriega, A., Grube, J. H., Talkad, J. & Sly, W. S. *J. Cell Biol.* **85**, 839-852 (1980).
14. Richardson, J. C. W. & Simmons, N. L. *FEBS Lett.* **105**, 201-204 (1979).
15. Caplan, M. J., Anderson, H. C., Palade, G. E. & Jamieson, J. D. *Cell* **46**, 623-631 (1986).
16. Huang, J. S., Huang, S. S. & Tang, J. T. *J. biol. Chem.* **254**, 11405-11417 (1979).
17. Hayman, E. G., Oldberg, A., Martin, G. R. & Ruoslahti, E. *J. Cell Biol.* **94**, 28-35 (1982).
18. Hay, E. D. in *Cell Biology of Extracellular Matrix* (ed. Hay, E. D.) 379-409 (Plenum, New York, 1981).
19. Timpl, R. *et al. J. biol. Chem.* **254**, 9933-9937 (1979).
20. Farquhar, M. G. In *Cell Biology of Extracellular Matrix* (ed. Hay, E. D.) 335-378 (Plenum, New York, 1983).
21. Madri, J. A., Roll, F. J., Furthmayr, H. & Foidart, J. M. *J. Cell Biol.* **86**, 682-687 (1980).
22. Stow, J. L., Sawada, H. & Farquhar, M. G. *Proc. natn. Acad. Sci. U.S.A.* **82**, 3296-3300 (1983).
23. Peters, B. P. *et al. J. biol. Chem.* **260**, 14732-14742 (1985).
24. Kelly, R. G. *Science* **230**, 25-32 (1985).
25. Matlin, K. S. *J. biol. Chem.* **261**, 15172-15178 (1986).
26. Laemmli, U. *Nature* **227**, 680-685 (1970).

Location of the ATPase site of myosin determined by three-dimensional electron microscopy

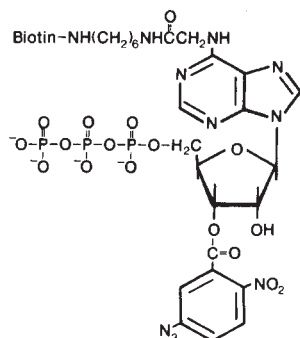
Makio Tokunaga*, Kazuo Sutoh†, Chikashi Toyoshima*‡ & Takeyuki Wakabayashi*

Departments of * Physics and † Biophysics and Biochemistry, Faculty of Science, University of Tokyo, Hongo, Bunkyo, Tokyo 113, Japan

Both ATP hydrolysis by myosin and the accompanying cyclic association-dissociation of actin and myosin are essential for muscle contraction. It is important for understanding the molecular mechanism of contraction to know the three-dimensional locations of the two major functional sites of myosin: the ATPase site¹ and the actin-binding site^{2,3}. We have determined the position of the ATPase site of myosin using three-dimensional image reconstruction from electron micrographs and site-specific labelling with the avidin-biotin system¹. The ATPase site is about 5 nm from the tip of the myosin head and is about 4 nm away from the actin-binding site of myosin. This is the first report of the three-dimensional location of an enzyme active site by electron microscopy.

Until now only X-ray crystallography has been able to show the three-dimensional location of functional sites of proteins. We report here an alternative approach¹³ combining the use of three-dimensional image reconstruction from electron microscopy⁴⁻¹¹ and site-specific labelling with the avidin-biotin system^{1,12} which we have used to locate the ATPase site of myosin.

Actin-tropomyosin-myosin subfragment-1-avidin complexes were prepared for electron microscopy. Instead of the biotinylated photo-affinity ADP analogue the ATP analogue was synthesized. Its structure was:



The ATP analogue was hydrolysed to the ADP analogue by incubation with myosin subfragment-1 (S1). The ADP analogue

‡ Present address: Department of Cell Biology, Stanford University School of Medicine, Stanford, California 94305, USA.

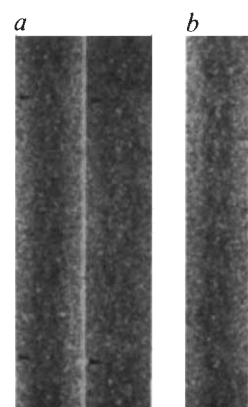


Fig. 1 *a*, Electron micrographs of the actin-TM filaments decorated with the purified S1-avidin complexes. *b*, The electron micrograph of the control specimen. The actin-TM filament was associated with intact S1 in the presence of ADP and avidin. Scale bar, 100 nm. *c*, Solid model of the three-dimensional image of actin-TM-S1-avidin reconstructed from electron micrographs. Note that there is a protrusion on the outer surface of S1 which has been assigned as avidin (Fig. 2). The boundary of S1 and avidin is shown by a broken line. Three molecules of S1 and avidin are seen on the near side. The filament axis is shown by a vertical line. Scale bar, 5 nm.

Methods. Myosin, S1 (ref. 31), actin and TM were prepared from rabbit skeletal muscle. Avidin was succinylated before use with succinic anhydride to decrease its positive charge¹⁴. Labelling of the ATPase site of S1 with avidin was performed as described previously¹. The mixture of succinylated avidin and S1 labelled with the ADP analogue was loaded on to a HABA (2-(4-hydroxyphenylazo)-benzoic acid)-conjugated column¹⁷. Bound proteins (the S1-avidin complex and free avidin) were eluted without denaturation with 0.1 mM biotin, 0.2 M NaCl, 20 mM imidazole-HCl, pH 7.0 and the concentrated eluate (Amicon) was loaded onto an HPLC gel filtration column (Asahi Pak GS-520) to separate the S1-avidin complex from free avidin. The purified S1-avidin complex (4 μ M) was mixed with actin-TM filament (1 μ M in actin). The mixture was dialysed against 10 mM sodium phosphate, 5 mM MgCl₂, pH 7.5 at 4 °C. For the control experiment *b*, a mixture of intact S1 (4 μ M), avidin (4 μ M) and actin-TM filament (1 μ M) in 10 mM sodium phosphate, 6 mM MgCl₂, 1 mM ADP, pH 7.5 was used. A JEOL 100CX electron microscope operating at 100 kV was used at a magnification of 32,400. The total electron dose was $<5 \text{ e}^- \text{ \AA}^{-2}$.

was then trapped into the ATPase site of S1 in the presence of vanadate and covalently crosslinked to the site by ultraviolet illumination. Complexes of S1 with avidin were formed for location in the electron microscope, avidin being a protein (relative molecular mass, M_r , 68,000) with a high affinity for biotin. Sutoh¹⁵ has recently shown that the site of covalent labelling is in a hydrophobic segment of ~10 residues adjacent to Trp 130, which has been located in the ATPase site by a

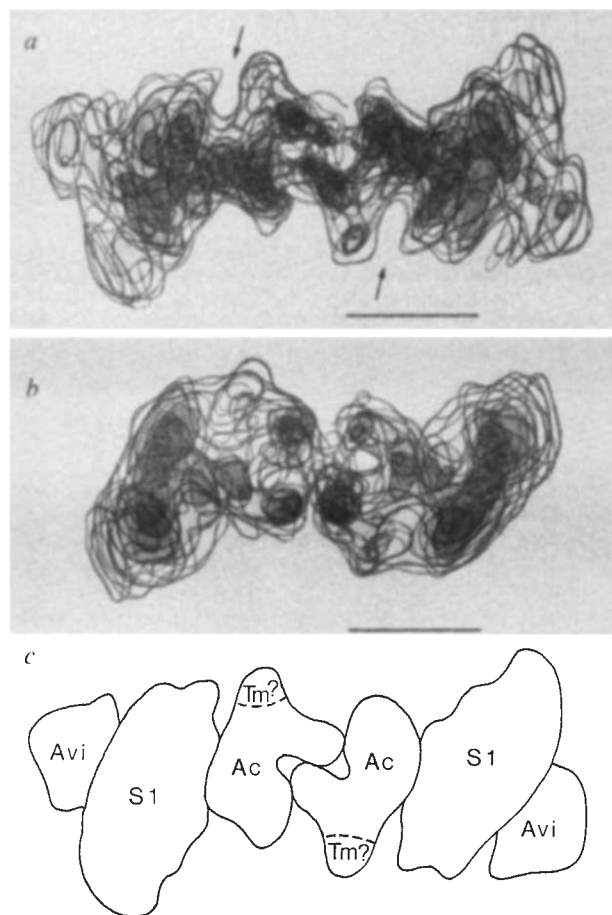


Fig. 2 *a*, End-on view of the transparent model of the three-dimensional image of the actin-TM-S1-avidin complex. Only two actin subunits are shown, together with the attached S1 molecules and tropomyosin strands associated with them. *b*, Reference image of the actin-TM-S1 complex¹¹. *c*, Diagram representing *a* as interpreted. The position of avidin was determined by comparison between *a* and *b* and by examining the difference map (*a* minus *b*, see text). Actin and S1 were assigned according to refs 2 and 3. Arrows in *a* indicate grooves between actin and S1. Scale, 5 nm. **Methods.** Three-dimensional image reconstruction was performed as in ref. 11 with slight modifications. The selection rule for helical symmetry was $l = -19n + 41m$. Because it was difficult to determine the polarity of the filaments by eye, the polarity of each filament was found by a fitting procedure in the Fourier space^{33,34}. The averaged intraparticle phase residual (parameter *Q*) (ref. 32) was 14.5° (12 pairs involved). The averaged interparticle phase residuals (parameter *P*) (ref. 34) was 59.5° (nine layer lines at 2.2 nm resolution). The images of eight filaments (only two filaments out of eight are shown in Fig. 1*a*) were averaged in the Fourier space. The layer-line data up to $1/2.6 \text{ nm}^{-1}$ (the 44th ($n=2$, $l=44$) layer-line) in the axial direction and up to $1/1.5 \text{ nm}^{-1}$ (the mean value) in the radial direction were used for the final Fourier-Bessel synthesis. The cut-off level of the three-dimensionally reconstructed image was selected so that the volume of S1 became equal to that of S1 in the reference image¹¹. The volume recoveries of avidin, S1 and actin-TM in the reconstructed image to the mass expected from the molecular weight were 30%, 70% and 75%, respectively.

photo-affinity labelling experiment¹⁶. Purification of native S1-avidin complexes from unbound avidin and S1 was essential for three-dimensional electron microscopy, because only 15% of the analogue trapped into the ATPase site was covalently crosslinked to S1. Native S1-avidin could be purified using an avidin affinity column and biotin as eluent. The purified S1-avidin complexes were incubated with actin tropomyosin (TM) filaments. At physiological ionic strength, there was only weak binding of S1-avidin complexes to filament but this increased

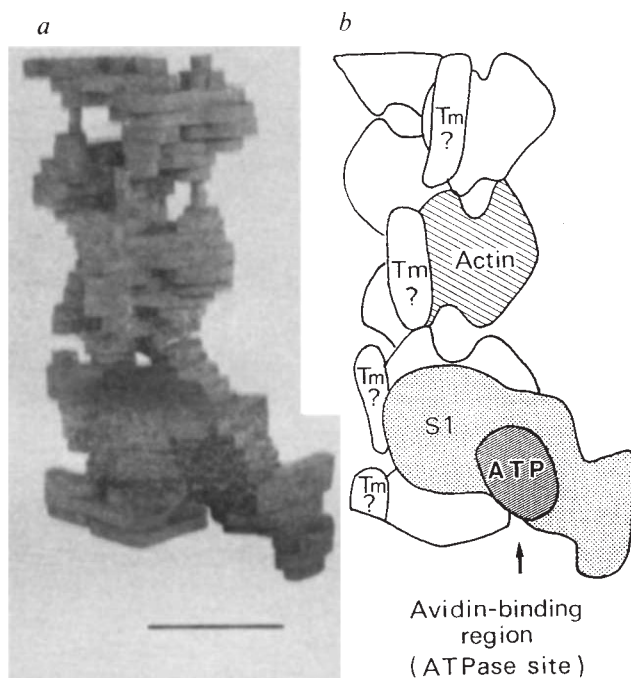


Fig. 3 *a*, ATPase site on the S1 molecule, which binds to actin-TM filament. The actin-TM filament and one S1 molecule were cut out of the solid model of the actin-TM-S1-avidin complex shown in Fig. 1*c*. The avidin-binding region is depicted in black. Scale bar, 5 nm. *b*, Schematic drawing of *a*. The ATPase site is within the avidin-binding region or very close to it.

to saturation at lower ionic strength; covalent crosslinking of the ADP analogue to S1 weakened the binding. Minimal dose electron micrographs of actin-TM-S1-avidin complexes negatively stained with 2% uranyl acetate (Fig. 1*a*) look different from those of actin-TM-S1 complexes (Fig. 1*b*: actin-TM-S1 in the presence of free avidin and ADP): the arrowhead appearance cannot be distinguished in the micrographs of the actin-TM-S1-avidin complexes but are recognizable in those of the actin-TM-S1 complexes.

The three-dimensional image of the actin-TM-S1-avidin complex was reconstructed from electron micrographs, using the helical symmetry¹¹. Eight images were averaged (only two micrographs of the eight are shown in Fig. 1*a*). The resolution was 1.5 nm radially and 2.6 nm axially. In the reconstructed image, there was a protrusion on the outer surface of S1 (Figs 1*c* and 2*a*). In the difference map (the density of actin-TM-S1-avidin minus that of actin-TM-S1 (ref. 11)), the largest peak located at this protrusion. The second largest peak was smaller (one-fifth in volume) and stretched over both the negative density region and the actin region. The other peaks were much smaller. Therefore, the protrusion was assigned to an avidin molecule bound to S1 (Fig. 2). The reconstructed images of actin-TM-S1-avidin and actin-TM-S1 showed similar shapes of S1 and actin (Figs 2 and 3). We noticed some differences, however: there were clear grooves between actin and S1 as indicated by the arrows in Fig. 2, and the discontinuous column-like structure that was the candidate for tropomyosin in the actin-TM-S1 filament² was clearer (Fig. 3). These differences could arise from the binding of ADP to S1, from the lower ionic strength of solution, or from a lower electron dose.

The size of the putative avidin protrusion on S1 was $\sim 3 \text{ nm} \times 4 \text{ nm} \times 4 \text{ nm}$ and was smaller than the size of avidin reported previously¹⁸ ($4 \text{ nm} \times 5.5 \text{ nm} \times 5.5 \text{ nm}$). The reason for this discrepancy could be the changing position of avidin. Because the four biotin-binding sites of an avidin molecule are away from the centre of the particle¹⁸, it is likely that bound avidin takes various orientations on S1, choosing one of the four binding

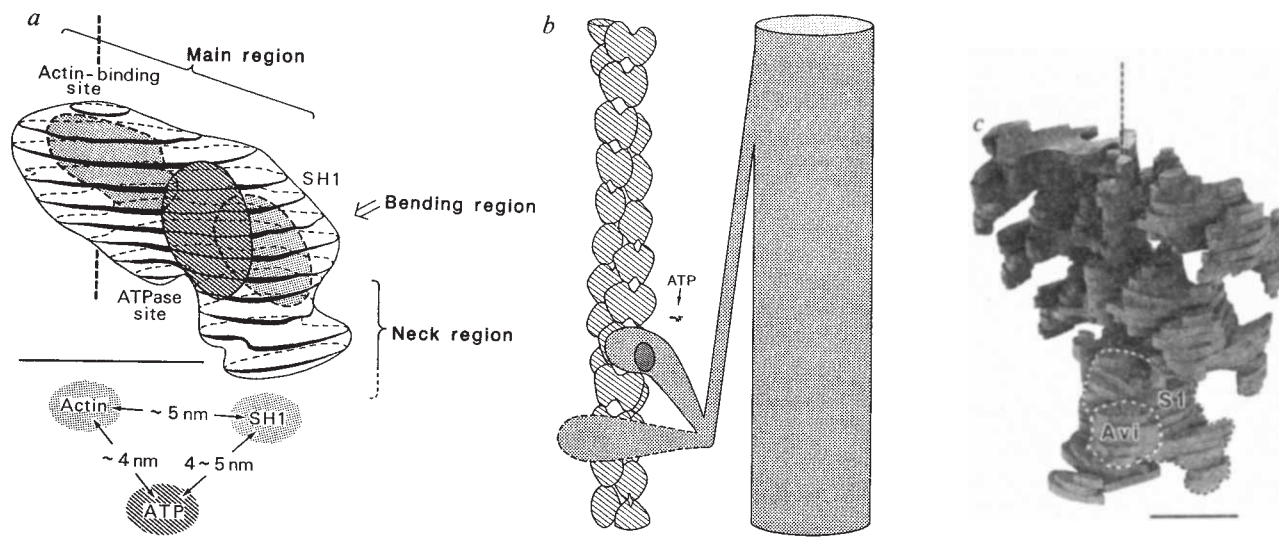


Fig. 4 *a*, Schematic diagram showing the three-dimensional location of the functional sites of the myosin head (S1), as viewed from the outside of the decorated filament. The ATPase site is on this side (hatched). The actin binding site^{2,3}, and SH1 (ref. 13) (the fast-reactive thiol of myosin) is on the opposite, or inner side (dotted). Distances between sites were measured from centre to centre. The direction of the axis of the actin filament to which this S1 molecule might be attached is shown by a vertical broken line. Note that the myosin head has a bent shape. The location of the ATPase site is near the bent region. The myosin head can be divided into two regions, a main region and a neck region^{7,11}. The main region is from the tip (the left end of the head in this figure) to the bent region (about 7 nm) and includes the ATPase site and the actin-binding site. The remaining region is the neck region. The density of the region near the head-rod junction of the reconstructed image of S1 has decreased owing to the absence of light chain 2 and to fluctuation of the structure. The neck region should therefore be longer than drawn here. Scale bar, 5 nm. *b*, Schematic diagram of a crossbridge, a thin filament (troponin is not shown), a thick filament and an ATP molecule, to show their size and the location of the ATPase site. The myosin head enclosed with solid line represents that in the rigor complex. The head enclosed with a broken line represents a possible configuration of the head not associated with actin. Scale bar, 10 nm.

sites. Another cause for position fluctuations might be the distance from the photoreactive site on the ribose ring to the biotin moiety linked to the adenine ring. As the length of the spacer arm of the ADP analogue is selected so that the adenine ring just reaches the surface of the avidin molecules¹⁸, the distance from the ribose ring to the surface of the avidin molecule should be about 1 nm.

To examine the avidin-binding region the avidin protuberance was cut out from the reconstructed image of the actin-TM-S1-avidin complex (Fig. 3). The avidin-binding region measured $\sim 2.5 \times 4$ nm. As avidin was bound to S1 through the biotin moiety of the ADP analogue crosslinked to the ATPase site, the ATPase site must be in, or very close to, the avidin-binding region.

The ATPase site we have located in the three-dimensional structure of S1 is on the opposite side of the actin-binding site^{2,3} (Fig. 4*a*). This position is favourable for ATP binding, being located on the outside of the actin-S1 complex. The distance between these two sites is ~ 4 nm (Fig. 4*a*). The actin-binding site is an allosteric site of the myosin ATPase, so communication with the ATPase site must be possible over the distance of separation of ~ 4 nm.

The distance from the ATPase site to the tip of S1 is ~ 5 nm along its long axis. Sutoh *et al.*¹ showed that the distance from the ATPase site to the head-rod junction of the myosin head was 14 nm. These two values are consistent, because the total length of the myosin head along its long axis is about 19 nm (refs 19–21). A fast-reactive thiol, SH1, has been found in the three-dimensional structure of S1 (ref. 13). The thiol was located on the same side as the actin-binding site (Fig. 4*a*) and the ATPase site was about 4–5 nm away from the SH1 thiol. Fluorescence energy transfer experiments²² have also shown that the two sites are 4 nm apart.

The myosin head has a bent shape (Fig. 4*a*) and can be divided into a main region and a neck region^{7,11} (Fig. 4*a*). The main region, whose axis is almost perpendicular to the helix axis^{6,7,11}, is the distal one-third of the head. It contains the ATPase site, as is shown in this study, and the actin-binding site. The neck

region, whose axis is roughly parallel to the helix axis¹¹, is the remaining two-thirds of the head, closer to the head-rod junction.

The tilting mechanism²³ of the whole myosin head (with some exceptions, for example, see ref. 24) has provided a working hypothesis for the molecular mechanism of muscle contraction, but experiment has so far failed to show such a tilting^{6,11,25–30}. Moreover, fluorescent or spin-labelled ATP or ADP bound to myosin did not change its angle to the axis of the actin filament during muscular contraction^{27,30}. It has been suggested¹¹ that muscle contracts by tilting the neck region of S1, but not the main region. This work shows that the ATPase site is in the main region near the bent region. Therefore, it is likely that the main region containing the actin-binding site and the ATPase site does not tilt during contraction.

The angle between the main region and the neck region of the myosin head could vary under some conditions (Fig. 4*b*). It is possible that during muscular contraction the neck region of the myosin head changes its orientation to the helix axis of the actin filament and that the orientation of the main region remains unchanged.

We thank Professor S. Ebashi and Dr F. Ebashi for support, Professors T. Yamazaki and T. Kamae for making a VAX-11/780 available, Dr L. A. Amos for providing a program package for processing large image arrays and Dr A. Tomioka and Mr T. Matsumoto for the use of their programs. This work was supported by the Ministry of Education, Science and Culture of Japan.

Received 20 February; revised 3 July; accepted 1 September 1987.

1. Sutoh, K., Yamamoto, K. & Wakabayashi, T. *Proc. natn. Acad. Sci. U.S.A.* **83**, 212–216 (1986).
2. Toyoshima, C. & Wakabayashi, T. *J. Biochem., Tokyo* **97**, 245–263 (1985).
3. Tomioka, A., Tokunaga, M. & Wakabayashi, T. *Proc. China-Japan Bilateral Symposium on Biophysics* 505–509 (China Academic, 1985).
4. DeRosier, D. J. & Klug, A. *Nature* **217**, 130–134 (1968).
5. Moore, P. B., Huxley, H. E. & DeRosier, D. J. *J. molec. Biol.* **50**, 279–295 (1970).
6. Toyoshima, C. & Wakabayashi, T. *J. Biochem., Tokyo* **86**, 1887–1890 (1979).
7. Wakabayashi, T. & Toyoshima, C. *J. Biochem., Tokyo* **90**, 683–701 (1981).
8. Taylor, K. A. & Amos, L. A. *J. molec. Biol.* **147**, 297–324 (1981).
9. Amos, L. A., Huxley, H. E., Holmes, K. C., Goody, R. S. & Taylor, K. A. *Nature* **299**, 467–469 (1982).

10. Vibert, P. & Craig, R. *J. molec. Biol.* **157**, 299-319 (1982).
11. Toyoshima, C. & Wakabayashi, T. *J. Biochem., Tokyo* **97**, 219-243 (1985).
12. Sutoh, K., Yamamoto, K. & Wakabayashi, T. *J. molec. Biol.* **178**, 323-339 (1984).
13. Toyoshima, C., Yamamoto, K., Sutoh, K., Sekine, T. & Wakabayashi, T. *11th YAMADA Conference, Abstr.* p. 38 (1985).
14. Holtzman, E., Wise, D., Wall, J. & Karlin, A. *Proc. natn. Acad. Sci. U.S.A.* **79**, 310-314 (1982).
15. Sutoh, K. *Biochemistry* (in the press).
16. Okamoto, Y. & Yount, R. G. *Proc. natn. Acad. Sci. U.S.A.* **82**, 1575-1579 (1985).
17. Bayer, E. & Wilchek, M. *Meth. Enzym.* **34**, 265-267 (1974).
18. Green, N. M., Konieczny, L., Toms, E. J. & Valentine, R. C. *Biochem. J.* **125**, 781-791 (1971).
19. Elliott, A. & Offer, G. *J. molec. Biol.* **123**, 505-519 (1978).
20. Takahashi, K. *J. Biochem., Tokyo* **83**, 905-908 (1978).
21. Walker, M., Knight, P. & Trinick, J. *J. molec. Biol.* **184**, 535-542 (1985).
22. Tao, T. & Lamkin, M. *Biochemistry* **20**, 5051-5055 (1981).
23. Huxley, H. E. *Science* **164**, 1356-1366 (1969).
24. Huxley, A. F. *J. Physiol., Lond.* **243**, 1-43 (1974).
25. Yagi, N., O'Brien, E. J. & Matsubara, I. *Biophys. J.* **33**, 121-138 (1981).
26. Huxley, H. E., Faruqi, A. R., Kress, M., Bordsas, J. & Koch, M. H. J. *J. molec. Biol.* **158**, 637-684 (1982).
27. Yanagida, T. *J. molec. Biol.* **146**, 539-560 (1981).
28. Cooke, R., Crowder, M. S. & Thomas, D. D. *Nature* **300**, 776-778 (1982).
29. Borejdo, J., Assulin, O., Ando, T. & Putnam, S. J. *J. molec. Biol.* **158**, 391-414 (1982).
30. Crowder, M., Wilkie, M. & Cooke, R. *Biophys. J.* **47**, 381a (1985).
31. Weeds, A. G. & Taylor, R. S. *Nature* **257**, 54-56 (1975).
32. DeRosier, D. J. & Moore, P. B. *J. molec. Biol.* **52**, 355-369 (1970).
33. Amos, L. A. & Klug, A. *J. molec. Biol.* **99**, 51-73 (1975).
34. Wakabayashi, T., Huxley, H. E., Amos, L. A. & Klug, A. *J. molec. Biol.* **93**, 477-497 (1975).

Preferential deletion of exons in Duchenne and Becker muscular dystrophies

S. M. Forrest*, G. S. Cross*, A. Speert†, D. Gardner-Medwin‡, J. Burn§ & K. E. Davies*

* Nuffield Department of Clinical Medicine, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, UK

† Department of Cell Differentiation, Academy of Sciences of DDR, Robert Rossle Strasse 10, 1115 Berlin, DDR

‡ Regional Neurological Centre, Newcastle General Hospital, Westgate Road, Newcastle-upon-Tyne NE4 6BE, UK

§ Department of Human Genetics, 19 Claremont Place, Newcastle-upon-Tyne NE1 4LP, UK

Duchenne and Becker muscular dystrophy (DMD and BMD) genes are located in Xp21 on the short arm of the X chromosome. DMD patients display a much more severe clinical course than BMD patients, and yet about 10% of cases of each have been reported to have deletions for parts of the gene. Using a complementary DNA subclone of the DMD gene we have screened 66 DMD and BMD patients who had not previously shown deletions with the probes then available. Fifteen patients have a deletion of this part of the gene, indicating a higher deletion frequency in this region (22%). Exons were deleted in five severely affected DMD patients and in ten BMD patients. Significantly, most of these deletions begin in the same region of the cDNA, which implies that there is a common mechanism for the generation of many of these mutations. An apparently identical deletion in one family gave classical BMD in two brothers (presenting in their teens) and only very mild muscle weakness in their 86-year-old great-great-uncle. Taking these data together with data using the probes previously published, we are able to detect deletions directly in 40% of our families requiring antenatal diagnosis or carrier detection.

Duchenne muscular dystrophy (DMD) is the most common X-linked disorder in man, affecting ~1 in 3,000 males (for review see ref. 1). Affected boys are usually wheelchair-bound by the age of 12 and die in their late teens. Becker muscular dystrophy (BMD) is characterized by a much milder clinical course in which the affected individuals stop walking at 12 years or later. Some BMD patients may have a normal lifespan. It was observed that females who suffer from the disease² have X;autosomal balanced translocations with one of the breakpoints within Xp21. Genetic studies using markers in this region have shown that DMD and BMD are both located in Xp21 (refs 3 and 4).

More recently, sequences have been isolated that are deleted in approximately 10% of both DMD and BMD patients (pERT87 (refs 5 and 6), XJ1.1 (ref. 7), HIP25 (ref. 8)). A candidate gene has been isolated^{9,10,16} and found to correspond to a large 14-kilobase (kb) messenger RNA which is expressed in both fetal and adult muscle.

Studies of deletions in patients with the available probes has demonstrated heterogeneity in the size of the deletions and no simple relationship between phenotype and genotype¹¹. But in view of the intragenic recombination rate of 5% (refs 12, 13), the deletions have proved important for accurate carrier detection and prenatal diagnosis. The deletion frequency, originally reported to be about 10%, is anticipated to be found to be higher once more of the gene sequence can be probed. We present here details of one particular region of coding sequence that is deleted in a highly significant number of patients and show that in most cases the deletion begins in the same region of the complementary DNA. These results are consistent with those of Koenig *et al.*¹⁰ who have also demonstrated a high deletion frequency in this region of the gene. In addition, we demonstrate that many of the BMD deletions begin with the same exon of the gene, whereas the DMD deletions are heterogeneous.

We isolated a 4.3-kb fetal muscle cDNA clone, Cf23, from the DMD region by walking from our previously reported adult cDNA clone^{14,15} in a randomly primed fetal muscle cDNA library. This original cDNA was isolated by screening an adult muscle oligo(dT)-primed library with an oligonucleotide from the published sequence in the pERT87 region⁹. Cf23 extends from within the pERT87 region to a region distal to JBir. The clone has a single *Pst*I site and detects at least 19 independent *Pst*I bands (which presumably correspond to 18 independent exons of the gene) in a Southern blot (Fig. 1). Many of these bands are doublets or triplets which makes it impractical to use this clone for deletion screening. We therefore subcloned the cDNA into two *Pst*I/*Eco*RI fragments, Cf23a (1.5 kb) and Cf23b (2.8 kb). The hybridization of the more telomeric Cf23a to patients with previously characterized deletions is shown in Fig. 2. Bands are deleted in patient 1477 who is otherwise deleted for JBir only (track 1). DNA from patient 346 (track 2) who is deleted from a region distal to, but not including, JBir, does not have a band at 12.5 kb but does hybridize to a fainter upper band which corresponds to a restriction fragment length polymorphism (RFLP). (JBir is a locus lying ~300 kb telomeric from the pERT87 locus.) The patient in track 3 is deleted in the pERT87 region but has JBir and all the exons of the cDNA.

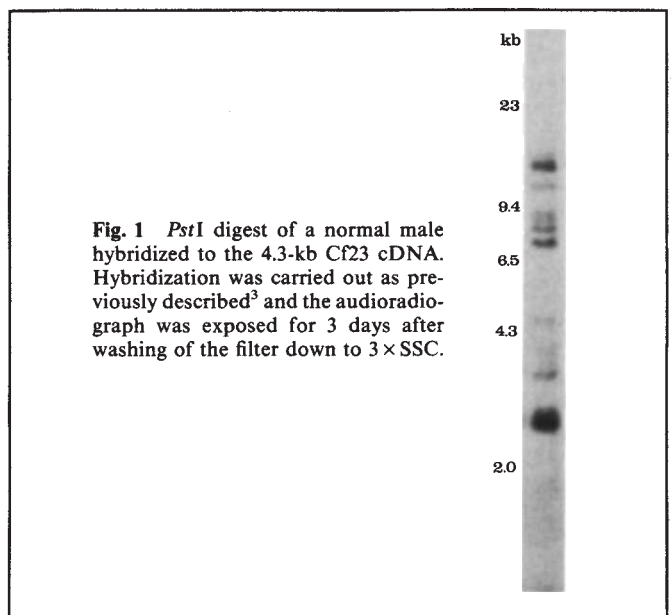


Fig. 1 *Pst*I digest of a normal male hybridized to the 4.3-kb Cf23 cDNA. Hybridization was carried out as previously described³ and the autoradiograph was exposed for 3 days after washing of the filter down to $3 \times \text{SSC}$.

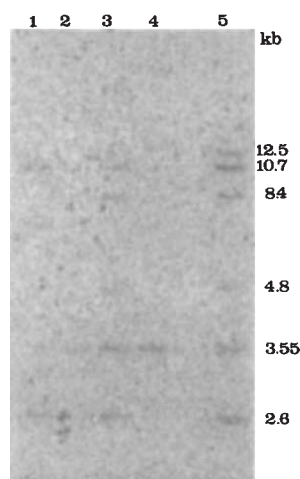


Fig. 2 *Pst*I digests of patients hybridized to Cf23a. Track 1, patient 1477 suffering from glycerol kinase deficiency and DMD; track 2, patient 346 suffering from adrenal hypoplasia, glycerol kinase deficiency and DMD; tracks 3 and 4, DMD males; track 5, male control. Conditions as in Fig. 1.

Track 4 is a DMD patient whose deletion extends from within the pERT87 locus to a position distal to JBir. He only has the 3.55-kb exon of the cDNA which must therefore be the most telomeric exon in Cf23a. This cDNA clone thus detects exons in the region of the JBir sequence.

Figure 3 shows the hybridization of Cf23a to patients with no previously detectable deletions. A deletion of the *Pst*I band at 3.55 kb can be seen in tracks 3, 6, 7 and 11. These are BMD patients: 28 unrelated BMD patients were included in this analysis and 10 possess deletions in the cDNA. These patients included those who were severely affected (in a wheelchair before the age of 20), and those who presented in adulthood with a much milder form of the disease. Nine of these ten patients have deletions beginning with the 3.55-kb band. To eliminate the possibility of a restriction fragment length polymorphism, 88 normal adults were screened in the same way and showed no detectable deletions. The deletion is also easily visualized by another restriction enzyme, *Taq*I (Fig. 4). None of our families had the correct structure to allow us to determine whether unequal crossing over was the mechanism for this common deletion.

The DMD patient in track 10 of Fig. 4 has an extra band without an apparent deletion. It is likely that this is an RFLP with the other allele being a doublet in the hybridization pattern because it is also observed in the normal controls. Five DMD patients out of 38 patients screened had deletions of some of the exons of Cf23a, two of them with deletions beginning at the 3.55-kb *Pst*I band. The DMD patients could have further exons deleted not detected on our blots by Cf23a. Alternatively, these patients may have an additional mutation which we have so far been unable to detect. We are currently analysing these deletions in the gene in more detail to find out how they might arise. One of the patients with a deletion beginning with the 3.55-kb band is our most severely affected boy, suffering rapid onset of muscle weakness and severe mental retardation. The deletion data do not show a consistent correlation between the region of the cDNA deleted and the degree of mental impairment.

It is interesting to note that the great-great-uncle in one BMD pedigree, now aged 86 years, has much milder muscle weakness than his descendants with classical BMD who presented with symptoms in their teens, and yet all three appear to have the same deletion. The great-great-uncle presented at the age of 64 because of frequent falling and muscular dystrophy was diagnosed. In retrospect he noted that he had never been as good

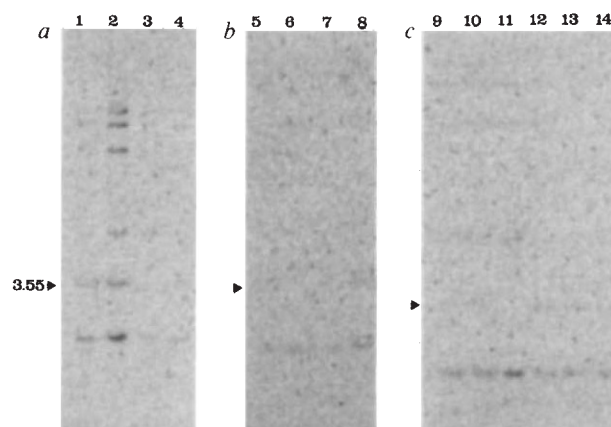


Fig. 3 a-c, Independent analyses of *Pst*I digests of BMD (tracks 1, 3, 4, 6, 7, 11 and 12) and DMD (tracks 5, 8, 9, 10, 13 and 14) males hybridized to Cf23a. Track 2 is a female with a clinical course typical of DMD.

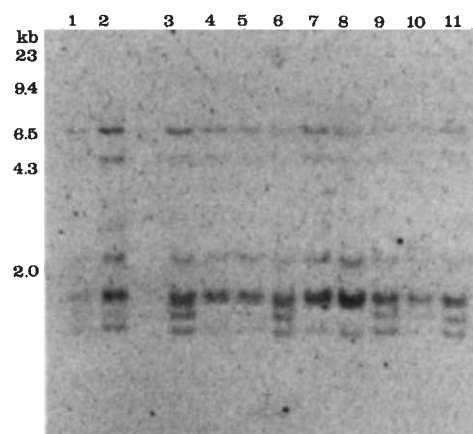


Fig. 4 *Taq*I digests of BMD (tracks 3, 4, 5, 7 and 8) and DMD (tracks 6, 9, 10 and 11) males. Tracks 1 and 2 are male and female controls respectively. Conditions as in Fig. 1.

at lifting or running as his colleagues but he was still employed as a miner and was able to do heavy manual work. At 70 years he had difficulty climbing stairs and became dependent on a wheelchair at 76 years. He was still able to move himself out of the chair until his legs became too weak at the age of 84. His descendants both presented in adolescence with classical BMD and their disease has shown a much more progressive course. For example, the older brother presented at 17 years complaining of difficulty walking upstairs even in early childhood. His muscle biopsy revealed severe chronic myopathy. The expression of the disease could thus depend on factors other than the deletion of a specific exon or there could be subtle changes not detected by this approach. This has important implications in the use of deletions for the prediction of the severity of the phenotype in a family.

The frequency of deletions observed in our patients is not increased by using cDNA probes lying 5' to Cf23a and covering 5.0 kb of cDNA sequence. We observe a deletion frequency of 10% (Smith, T. J. *et al.*, unpublished observations) and Koenig *et al.*¹⁰ observed a frequency of 14.5% in this 5' region. The data presented here suggest that a much higher frequency of deletions can be detected if specific regions of the gene closer to the 3' end are used as probes. The fact that 15 deletions were found in 66 patients make this cDNA sequence very useful for diagnostic purposes. Combining the use of this probe together

with pERT87, XJ1.1 and HIP25 (ref. 17), we are able to detect deletions in 40% of patients.

We thank Tracey Flint for expert technical assistance and many clinical colleagues for patient material, particularly Drs Bunday, McDermott, Super, Clarke, Bates, Hudgson, Spiegler and Metzka. This work was supported by the Muscular Dystrophy Group of Great Britain, the MRC, and the Muscular Dystrophy Association, USA. A.S. was funded by a Wellcome Trust Travelling Fellowship. We thank Rachel Kitt for assistance in preparing the manuscript. S.F. is an Ackman Scholar.

Received 4 August; accepted 7 September 1987.

- Emery, A. E. H. *Oxford Monographs on Medical Genetics, No. 15: Duchenne Muscular Dystrophy* (Oxford University Press, 1987).
- Lindenbaum, R. H., Clarke, G., Patel, C., Moncrieff, M. & Hughes, J. T. *J. med. Genet.* **16**, 389-392 (1979).
- Davies, K. E. *et al. Nucleic Acids Res.* **11**, 2303-2312 (1983).
- Kingston, H. M., Sarfarazi, M., Thomas, N. S. T. & Harper, P. S. *Hum. Genet.* **67**, 6-17 (1984).
- Kunkel, L. M., Monaco, A. P., Middlesworth, W., Ochs, H. D. & Latt, S. A. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4778-4782 (1985).
- Kunkel, L. M. *et al. Nature* **322**, 73-77 (1986).
- Ray, P. N. *et al. Nature* **318**, 672-675 (1985).
- Smith, T. J. *et al. Nucleic Acids Res.* **15**, 2167-2174 (1987).
- Monaco, A. P. *et al. Nature* **323**, 646-650 (1986).
- Koenig, M. *et al. Cell* **50**, 509-517 (1987).
- J. med. Genet.* December issue (1986).
- Monaco, A. P. & Kunkel, L. M. *Trends Genet.* **3**, 33-37 (1987).
- Robertson, M. *Nature* **327**, 372-373 (1987).
- Brockdorff, N. *et al. Nature* **328**, 166-168 (1987).
- Cross, G. S. *et al. EMBO J.* (in the press).
- Burghes, A. H. M. *et al. Nature* **328**, 434-437 (1987).
- Davies, K. E. *et al. J. med. Genet.* (in the press).

Direct detection of more than 50% of the Duchenne muscular dystrophy mutations by field inversion gels

J. T. den Dunnen, E. Bakker, E. G. Klein Breteler, P. L. Pearson & G. J. B. van Ommen

Department of Human Genetics, State University Leiden, Wassenaarseweg 72, 2333 AL Leiden, Netherlands

Duchenne muscular dystrophy (DMD) is an X-linked disorder affecting about 1 in 3,500 males^{1,2}. It is allelic with the milder Becker muscular dystrophy. The biochemical basis for both diseases is unknown and no effective treatment is available. Long-range physical mapping has shown that the *DMD* gene, localized in Xp21, is extremely large²⁻⁶, exceeding 2 million base pairs⁷. Until now, carrier detection and prenatal diagnosis has involved the use of linked restriction fragment length polymorphism markers^{8,9} which detect muscular dystrophy-associated deletions in about 10% of the cases¹⁰⁻¹². Field inversion gel electrophoresis (FIGE) allows the detection of structural rearrangements in 21 out of 39 of the DMD patients studied (54%), of which 14 (65%) were not detected by conventional methods. Large deletions seem to make up a much higher fraction of the DMD mutations than so far indicated by other methods. A region prone to deletion was located in the distal half of the gene. FIGE analysis could provide a valuable extension of information for carrier detection and prenatal diagnosis. The technique should be generally applicable to the study of diseases involving structural chromosomal rearrangements.

After the finding that pERT87 probes detect deletions in 7% of the DMD patients¹⁰⁻¹², additional probes were isolated which raised the deletion fraction to between 11 and 17% (ref. 13 and Table 1). The use of rare cutting restriction enzymes in combination with pulsed field gradient (PFG) (ref. 14) or field inversion gel electrophoresis (FIGE) (ref. 15) should allow the detection of chromosomal rearrangements (including inversions and translocations) at larger distances from a given probe than permitted by conventional techniques. Also more information

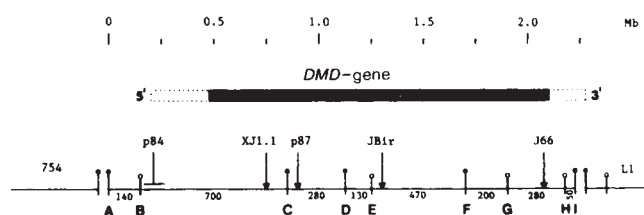


Fig. 1 Schematic map of the DMD region. Details of the map for the restriction enzyme *SfiI* are from van Ommen *et al.*⁷ Filled circles, *SfiI* sites; open circles, partially digested sites. Each site has a letter identification (A-I) to facilitate discussion; fragment sizes are indicated in kb. Probes mentioned in the text are shown on the map. The region including the putative *DMD* gene and its orientation are indicated at the top, with distances given in millions of basepairs (Mb). Controlled *SfiI* partial digestion gives increased fragment sizes, which improves the sensitivity of the analysis. In this way 2.2 Mb of DNA, including the *DMD* gene, are covered using only four probes; XJ1-1 (ref. 19), pERT87-30 (ref. 12), JBIr (ref. 12) and J66 (ref. 7). In later experiments p84-10 (ref. 12) was used instead of XJ1-1 because it gave more information (is less often deleted itself).

will be available about the nature, location and extension of these rearrangements.

On the basis of the megabase maps of the DMD region (refs 4-7; also Fig. 1) FIGE analysis was performed on *SfiI* restriction enzyme digests. The absence of *SfiI* restriction fragment length polymorphisms (RFLPs) and other common variations unrelated to DMD was verified by testing a panel of unrelated normal individuals and by comparing DNA isolated from lymphocytes transformed by Epstein-Barr virus (EBV) with leukocyte DNA (Table 1). The results obtained with 39 DMD patients are summarized in Table 1. FIGE analysis showed 21 rearrangements (54%), 20 deletions and one duplication. Only seven of the deletions were also detected by conventional analysis, an example being patient DL62.5 (Fig. 2a). Hybridization with probe JBIr gives a normal signal, XJ1-1 gives no signal (not shown) and p84-10 and p87-30 detect an altered 750 kilobase (kb) *SfiI* fragment. Therefore this patient has a 370-kb deletion which spans the *SfiI*/C site and the XJ-region (Fig. 1). In contrast, conventional analysis of the DNA from patient DL23.5 showed no deletion. On a FIGE blot, JBIr detects a smaller

Table 1 A summary of results from 39 DMD patients using conventional or FIGE analysis

Method	Probe XJ1-1	Probe p87	Probe JBIr	Probe J66	Fraction	%
Southern blot*	4/43	15/104	5/74	0/86	18/104	17
FIGE—control	0/31	0/46	0/48	0/37	0/≥31	0
FIGE—DMD	+	+	+	+	16	41
	[]	+	+	+	1	3
	[]	+	+	+	3	8
	[]	—	[]	+	2	5
	+	[]	—	[]	1	3
	+	+	[]†	+	2	5
	+	+	[]	[]	6	15
	+	+	+	[]	6‡	15
Total FIGE					21	54 (§‡)

Values given (*d/n*) show the number of abnormalities detected (*d*) after screening *n* patients (or *n* X chromosomes in FIGE controls). For FIGE analysis, each type of abnormality detected is shown separately: (+) for a normal *SfiI* fragment, (—) for a deletion and square brackets for an altered fragment.

* Figures given are not corrected for deletions spanning more than one probe.

† Including a duplication.

‡ Including a small deletion (~20 kb), lying at the detection limit of our analysis.

§ 23/39 or 59%, including two cases requiring further study.

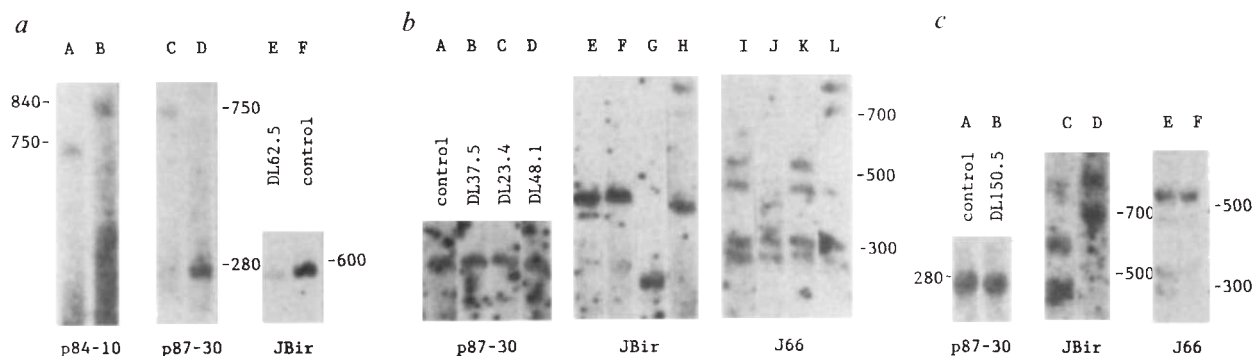


Fig. 2 Hybridizations of Xp21 probes to fragments of *SfiI*-digested DNA of several DMD patients separated by field inversion gel electrophoresis. *a*, Hybridizations on the same filter of the indicated probes to DNA from a control (lanes B, D and F) and patient DL62.5 (lanes A, C and E). Owing to partial digestion, the *SfiI*/B for p84-10) and *SfiI*/E (for JBir) sites are not cleaved, resulting in the fragment sizes indicated (in kb). *b*, Hybridizations of the probes indicated to DNA from a control person and three DMD patients. *SfiI* digestions are more extensive than in *a*, giving in the control lanes fragment sizes of 280 kb for pERT87-30, 470 kb (and a faint 600-kb band not visible on this exposure) for JBir and 530/480/330/280 kb for J66. Size markers apply to the complete panel. *c*, Hybridizations of pERT87-30, JBir and J66 to DNA from patient DL150.5 and a control.

Methods. DNA was isolated from EBV-transformed leukocytes as described previously²⁰. Partial *SfiI* digestions were performed for 6 h at 50–55 °C (in 0.5% LGT-agarose) using 25 units of enzyme (Biolabs) and DNA isolated from 0.75 million cells²⁰. Samples were loaded on 1% agarose gels. Standard FIGE gels were run at 7 V/cm⁻¹ in 0.5 × TBE (45 mM Tris/45 mM boric acid/0.5 mM EDTA, pH 8.3) at 18 °C. A standard run of 16 h was divided into four equal cycles of 4 h. Within each cycle the forward phase increased from 1–55 sec and the reverse phase lasted 33% of the forward phase. Switch time increased non-linearly: at 40% of a cycle, 50% of the switch time increase was completed. Gels were blotted onto HYBOND membranes (Amersham) and hybridized for 16 h in 0.25 M sodium phosphate/0.25 M NaCl/7% SDS/1 mM EDTA/10% PEG-6000 at 65 °C. Filters were washed from 2 × SSC/0.1% SDS down to 0.3 × SSC/0.1% SDS at 65 °C. Autoradiography was for 1–5 days at –70 °C with intensifying screens.

fragment (260 kb), but the p87 and J66 fragments are normal (Fig. 2b). Therefore this patient has a 210-kb deletion distally in the *SfiI*/E–F fragment.

Two predominant classes of chromosomal aberrations are represented by patients DL48.1 and DL37.5 (Fig. 2b). DL48.1 has altered fragments, detectable with both JBir and J66, and other fragments specific for each probe. This can be explained by a 210-kb deletion removing the *SfiI*/F-site (Fig. 1). The resistance to complete digestion of the *SfiI*/G site causes large, novel fragments to be seen with both JBir and J66, whereas the products of more extended digestion, *SfiI*/G–H and *SfiI*/G–I, are not changed. In patient DL37.5 the 530/480-kb J66 fragments have altered mobilities of 400/350 kb, and the JBir fragment is unaltered. This patient therefore has a 130-kb deletion in the 200-kb *SfiI*/F–G fragment. Among the 21 rearrangements found, 12 (64%) fall into one of these two classes (Table 1). The high proportion of distal mutations is not a result of a sampling bias, as the number of proximal mutations in this patient set (7/39 cases, or 18%) is coincidentally higher than that in our total family material (15/104 or 14%, Table 1). Apparently the distal area of the *DMD* gene contains a deletion-prone mutational hotspot, a fact substantiated by standard RFLP results in which the mutation is usually distal to pERT87 (ref. 16).

Figure 2c shows a duplication found in patient DL150.5. JBir detects a larger *SfiI*-D/F doublet of 690/830 kb with a twofold autoradiographic intensity. Both flanking probes give normal hybridization signals, indicating that there was a 220-kb duplication (including JBir), confirmed by conventional analysis (not shown). The FIGE result is easier to interpret and allows direct assessment of the length of the duplication. Conventional analysis of this family provided insufficient data to establish the carrier status of DL150.4 (Fig. 3) FIGE analysis showed that the 230 kb duplication found in patient DL150.5 (Fig. 2c) was also present in his mother, implying that she was a carrier. DL150.4 did not show the duplication and is therefore not a carrier.

FIGE-analysis thus provides a valuable extension to the diagnosis of X-linked muscular dystrophy. In 54% of the cases, chromosomal aberrations are located directly and their nature is identified with great precision. The remaining cases could

result from point mutations or deletions¹⁰ smaller than our operational detection limit of 30 kb. FIGE analysis should be useful for prenatal diagnosis as the results take no longer than conventional techniques, particularly because the aberration can be identified in advance from the other family members.

The cloning of the entire *DMD* complementary DNA has recently been published¹⁷. The use of the cDNA to detect deletions^{17,18} would be a straightforward approach to diagnosis. The results confirm our finding that over 50% of all DMD mutations are deletions¹⁷. But the highly complex patterns caused by the large number of widely spread exons require subdivision of the cDNA into many small segments¹⁷. A great number of parallel or subsequent analyses thus become necessary to screen the entire gene. The division of the *DMD* gene into a limited number of well-separated FIGE fragments should offer a convenient solution to this problem. Notably, detection of deletion (or duplication) carriers by abnormal junction fragments, with the normal fragment serving as internal control, is more reliable than the quantitative interpretation of hybridization signals which is normally required to distinguish these mutations in

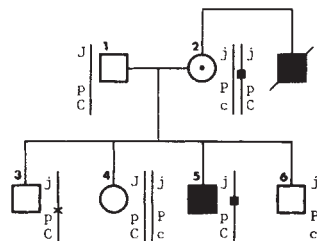


Fig. 3 Pedigree of a Dutch family (DL150) with a familial case of DMD. The mother (person 2), an obligate carrier, has one affected (person 5) and two non-affected sons (persons 3 and 6); her daughter (person 4) came for counselling. Informative probes were C7 (J/j), p87-1 (P/p) and 754 (C/c); not informative were the distal markers pD2, 99-6 and B24 (ref. 12). The cross (DL150.3) indicates a recombination and the black square indicates the duplication detected with JBir in a FIGE analysis. RFLPs were found with the probes described¹³. DNA for FIGE analysis was isolated from white blood cells in LGT-agarose blocks²⁰.

carrier females. Whichever technique or combination of techniques is favoured for clinical diagnosis, our results show that FIGE analysis greatly facilitates the study of the contribution of structural chromosomal rearrangements to hereditary and acquired diseases.

We thank M. Simons, M. Rijnkels, E. Bonten, L. de Lange, R. D. Runia and L. Gerrese for their expert technical assistance, R. G. Worton and L. M. Kunkel for providing probes, and M. H. Breuning and M. C. Wapenaar for helpful discussions. J.T.dD. was supported by the Muscular Dystrophy Group of Great Britain and this work was further supported by grants from MEDIGON, ZWO and the Dutch Prevention Fund.

Received 13 July; accepted 2 September 1987.

1. Moser, H. *Hum. Genet.* **66**, 17-40 (1984).
2. Monaco, A. P. & Kunkel, L. M. *Trends Genet.* **3**, 33-37 (1987).
3. Monaco, A. P. *et al. Nature* **323**, 646-650 (1986).
4. Burmeister, M. & Lehrach, H. *Nature* **324**, 482-485 (1986).
5. Van Ommen, G. J. B. *et al. Cell* **47**, 499-504 (1986).
6. Kenrick, S., Patterson, M., Speer, A., Fischbeck, K. & Davies, K. *Cell* **48**, 351-357 (1987).
7. van Ommen, G. J. B. *et al. Genomics* (submitted).
8. Wieacker, P., Davies, K. E., Pearson, P. L. & Ropers, H.-H. *Lancet* **ii**, 1325-1327 (1983).
9. Bakker, E. *et al. Lancet* **i**, 655-658 (1985).
10. Monaco, A. P. *et al. Nature* **316**, 842-845 (1985).
11. Kunkel, L. M. *et al. Nature* **322**, 73-77 (1986).
12. Monaco, A. P., Bertelson, C. J., Colletti-Feener, C. & Kunkel, L. M. *Hum. Genet.* **75**, 221-227 (1987).
13. Bakker, E. *et al. J. med. Genet.* **23**, 573-580 (1986).
14. Carle, G. R. & Olson, M. V. *Nucleic Acids Res.* **12**, 5647-5664 (1984).
15. Carle, G. R., Frank, M. & Olson, M. V. *Science* **232**, 65-68 (1986).
16. Fischbeck, K. H. *et al. Lancet* **ii**, 104 (1986).
17. Koenig, M. *et al. Cell* **50**, 509-517 (1987).
18. Burghes, A. H. M. *et al. Nature* **328**, 434-437 (1987).
19. Ray, P. N. *et al. Nature* **318**, 672-675 (1985).
20. van Ommen, G. J. B. & Verkerk, J. M. K. in *Human Genetic Disease, A Practical Approach* (ed. Davies, K. E.) 113-133 (IRL, Oxford, 1986).

The retinoblastoma susceptibility gene encodes a nuclear phosphoprotein associated with DNA binding activity

Wen-Hwa Lee, Jin-Yuh Shew, Frank D. Hong,
Theodore W. Sery*, Larry A. Donoso*,
Lih-Juan Young, Robert Bookstein & Eva Y.-H. P. Lee

Experimental Pathology Program and Center for Molecular Genetics,
Department of Pathology M-012, University of California San Diego,
School of Medicine, La Jolla, California, 92093, USA

*Wills Eye Hospital, Ninth and Walnut Street, Philadelphia,
Pennsylvania 19107, USA

The human gene (*RB*) that determines susceptibility to hereditary retinoblastoma has been identified recently by molecular genetic techniques^{1,2,14}. Previous results indicate that complete inactivation of the *RB* gene is required for tumour formation. As a 'cancer suppressor' gene, *RB* thus functions in a manner opposite to that of most other oncogenes. Sequence analysis of *RB* complementary DNA clones demonstrated a long open reading frame encoding a hypothetical protein with features suggestive of a DNA-binding function². To further substantiate and identify the *RB* protein, we have prepared rabbit antisera against a trypE-*RB* fusion protein. The purified anti-*RB* IgG immunoprecipitates a protein doublet with apparent relative molecular mass (M_r) of 110,000-114,000. The specific protein(s) are present in all cell lines expressing normal *RB* mRNA, but are not detected in five retinoblastoma cell lines examined. The *RB* protein can be metabolically labelled with ³²P-phosphoric acid, indicating that it is a phosphoprotein. Biochemical fractionation and immunofluorescence studies demonstrate that the majority of the protein is located within the nucleus. Furthermore, the protein can be retained by and eluted from DNA-cellulose columns, suggesting that it is associated with DNA binding activity. Taken together, these results imply that the *RB* gene product may function in regulating other genes within the cell.

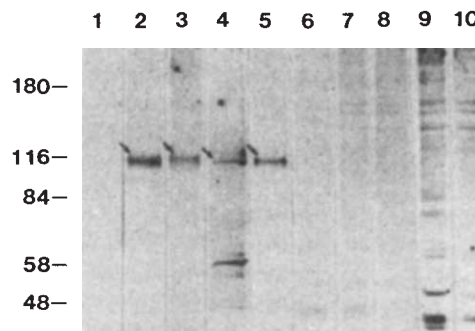


Fig. 1 Identification of *RB* proteins by immunoprecipitation with rabbit anti-*RB* IgG. Human cells such as neuroblastoma LAN-1 (lanes 1 and 2), Alexander hepatoma (lane 3), osteosarcoma U2OS (lane 4), normal fibroblasts (lane 5), and five retinoblastomas (lanes 6-10) were labelled with ³⁵S-methionine and immunoprecipitated with preimmune rabbit IgG (lane 1) or rabbit anti-*RB* IgG (lanes 2-10). The immunoprecipitates were analysed by 7.5% SDS-polyacrylamide gel electrophoresis and autoradiographed. The specific proteins (arrows) were not present among nonspecific background bands in lanes 6-10 despite prolonged exposure. Markers at left, M_r in thousands.

Methods. A protocol for immunoprecipitation was adapted as described⁵. Briefly, about 1×10^6 cells were labelled with ³⁵S-methionine ($150 \mu\text{Ci ml}^{-1}$) for 3 h. Cell extracts were prepared in lysis buffer containing 25 mM Tris-HCl (pH 7.4), 50 mM NaCl, 0.2% Nonidet P-40 (NP-40), 0.5% deoxycholate and 200 units ml^{-1} of aprotinin. 0.02% SDS was also added at the beginning of lysis. The lysates were clarified by centrifugation (4°C , 20,000g) for 15 min. Immunoprecipitation was carried out with 5 μl of preimmune rabbit antisera, followed by absorption to formalin-fixed *Staphylococcus aureus* (Enzyme Center). To supernatants were added 10 μl of anti-*RB* IgG ($100 \mu\text{g ml}^{-1}$) (or preimmune sera for control) and protein-A-sepharose beads (Sigma). Immunoprecipitates were subsequently washed with lysis buffer, then with 1 M NaCl in lysis buffer, then with 0.15 M NaCl in lysis buffer, and finally with lysis buffer to remove nonspecifically bound proteins. The immunoprecipitated proteins were analysed by 7.5% SDS-polyacrylamide gel electrophoresis. Gels of ³⁵S-labelled proteins were fluorographed at -70°C after impregnation with acetic acid-based 2,5-diphenyloxazole.

Using our hypothetical protein sequence data², we constructed three recombinant plasmids that express trpE fusion proteins³. Three fragments (of 0.9 kilobases (kb), 0.7 kb and 1.8 kb) obtained from *RB* cDNA clones² were fused in-frame into pATH vectors⁴ to produce recombinant constructs pATH-0.9, pATH-0.7 and pATH-1.8 respectively. pATH-0.9 and pATH-1.8 produced no protein in *Escherichia coli*, not even trpE itself. Because the *RB* sequence contains many hypothetical protease cleavage sites, this is probably the result of instability of the fusion protein. However, pATH-0.7 expressed a fusion protein with M_r 57,000 (57 K), of which 37K was trpE- and 20 K was *RB*-derived. Large quantities of fusion protein were prepared and purified by preparative SDS-polyacrylamide gel electrophoresis (PAGE). About 6 mg of protein was recovered. Two rabbits were immunized with the protein following a standard protocol⁵. Two months later, both rabbits produced high titres of antibodies that reacted with both trpE and the fusion protein. To enrich for antibodies recognizing only *RB* determinants, two affinity columns were prepared, one with trpE protein and the other with the fusion protein. Antisera were passed through the fusion protein/Sepharose column and eluted with glycine buffer (pH 2.3). The eluate was neutralized and passed through the trpE column several times to remove antibody directed against trpE. This purified anti-*RB* antibody was used in all subsequent experiments.

To identify the *RB* protein, we selected several human cell lines known to have either normal or altered *RB* gene expression. Control cell lines containing normal *RB* messenger RNA included LAN-1 (neuroblastoma)⁶, Alexander (hepatoma),

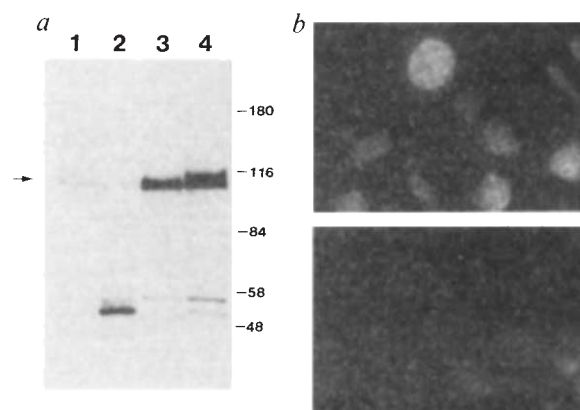


Fig. 4 Localization of the *RB* protein. *a*, Some ^{35}S -methionine labelled LAN-1 cells (lane 4) were fractionated into membrane (lane 1), cytoplasm (lane 2) and nucleus (lane 3) and the *RB* protein (arrow) was immunoprecipitated with anti-*RB* IgG. The immunoprecipitates were then analysed by SDS-PAGE as described for Fig. 1. *b*, Immunofluorescence studies of *RB* protein localization within osteosarcoma cell line U2OS. Cells reacted with (top) anti-*RB* IgG and (below) preimmune rabbit IgG. Most fluorescence was found within the nucleus.

Methods. The protocol for cell fractionation was adapted from that of Radke *et al.*¹¹. Two to five 100-mm plates containing a total of 2.0×10^7 LAN-1 cells were metabolically labelled with ^{35}S -methionine for 2–3 h before use. All subsequent procedures were performed at 4 °C. Cells were rinsed twice with phosphate-buffered saline (PBS), scraped into PBS, and pelleted for 5 min at 375 g in a tabletop centrifuge. Half the cells were resuspended in lysis buffer for the whole cell lysate. The remaining cells were rinsed once in hypotonic RSB buffer (10 mM HEPES pH 6.2, 10 mM NaCl, 1.5 mM MgCl_2 , 200 units ml^{-1} aprotinin) and then resuspended in RSB. The cell suspension was homogenized by 20 strokes in a tight fitting Dounce homogenizer, and volume was adjusted to exactly 3 ml with RSB buffer. The homogenate was centrifuged at 1,500 r.p.m. in a Sorvall HB4 rotor (375 g) for 10 min, and the pellet was resuspended in lysis buffer to generate the nuclear fraction. The supernatant was centrifuged in thick-walled polyallomer tubes in a Beckman SW50.1 rotor at 35,000 r.p.m. (150,000 g) for 90 min. The pellet was resuspended in lysis buffer to generate the membrane fraction, whereas the supernatant was adjusted to 1× lysis buffer concentration to produce the cytoplasmic fraction. All four fractions were then assayed for *RB* protein content as in Fig. 1. In *b*, immunofluorescent staining was carried out as described¹¹. Briefly, about 10^4 U2OS cells were seeded onto 12-mm glass cover slips and used 18 h later for immunofluorescent staining. The cells were washed once with PBS buffer and fixed with cold acetone for 10 min at room temperature (RT). Fixed and permeabilized cells were hydrated in PBS for 1–2 min. Each cover slip was incubated with 200 μl of rabbit anti-*RB* IgG (1:20 dilution) or preimmune serum in a moist chamber for 45 min at RT. After three washes in PBS, the cover slips were incubated with 200 μl of rhodamine-conjugated goat anti-rabbit IgG (25 $\mu\text{g ml}^{-1}$) (Sigma) for 45 min at RT. The cover slips were again washed extensively in PBS and viewed with a Zeiss photomicroscope III.

G+C rich region that translates into an unusual cluster of Ala and Pro residues (Fig. 2), it remains to be determined whether the first or second Met is the authentic initiation site.

Discrepancies between actual and apparent M_r on SDS-PAGE also may be explained by secondary protein modifications. There are several potential *N*-linked glycosylation sites in the predicted amino-acid sequence². When LAN-1 cells were grown in medium supplemented with ^{14}C -galactose or ^3H -glucosamine, labelled *RB* protein was not detected despite prolonged autoradiography (data not shown). In addition, digestion of ^{35}S -labelled *RB* protein by endoglycosidase H (ref. 8) did not result in a reduction of apparent M_r (Fig. 3). To test for protein phosphorylation, LAN-1 cells were metabolically labelled with ^{32}P -phosphoric acid. Immunoprecipitated protein ran as a single band with M_r identical to that of ^{35}S -labelled

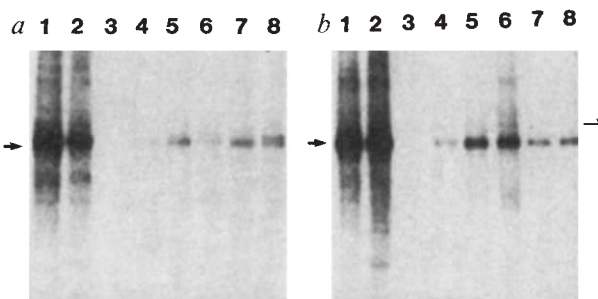


Fig. 5 Column chromatography of pp110^{RB} on single-stranded and double-stranded DNA cellulose. Protein lysates of LAN-1 cells metabolically labelled with ^{32}P -phosphoric acid were passed through single stranded (*a*) and double-stranded (*b*) DNA cellulose columns and eluted with 0.05, 0.1, 0.2, 0.3, 0.5 and 1.0 M NaCl (lanes 3–8, respectively) stepwise. The eluates containing the *RB* protein (arrow) were then immunoprecipitated with rabbit anti-*RB* IgG. Lanes 1 and 2 contain the whole cell lysates and flow-through fraction immunoprecipitated with anti-*RB* IgG respectively.

Methods. LAN-1 cells labelled with ^{32}P phosphoric acid for 3 h were lysed in lysis buffer and clarified as described in Fig. 1. The supernatant was diluted 10-fold with loading buffer (1 mM dithiothreitol (DTT), 0.5% NP-40, 10 mM potassium phosphate pH 6.2, 10% glycerol, aprotinin 200 units ml^{-1}). The diluted extract was then applied to calf thymus DNA-cellulose columns (Pharmacia)¹² equilibrated with loading buffer containing 50 mM NaCl. After binding had been allowed to occur for 40 min, the column was washed with 5 ml of loading buffer and then eluted with buffer containing 1 mM DTT, 10 mM Tris-HCl, pH 8.0 with increasing NaCl concentration from 0.05 to 1.0 M. fractions were then analysed by immunoprecipitation as described in Fig. 1.

RB protein (Fig. 3), indicating that the *RB* protein is a phosphoprotein which we shall call pp110^{RB} .

The *RB* gene is detectable in the DNA of other vertebrates. When cells from quail (QT6), mouse (NIH 3T3), rat (Rat-2) and monkey (Cos) were labelled with ^{32}P -phosphoric acid and their proteins immunoprecipitated with anti-*RB* immunoglobulin G (IgG), antigenically related proteins of between 108 and 128K were detected in all of them (data not shown).

It has been proposed that the *RB* gene product acts as a regulator of other genes^{9,10}. Moreover, from features of the predicted amino-acid sequence², we suggested that the *RB* protein may have DNA-binding activity. A prerequisite for these functions is localization within the nucleus. We therefore undertook biochemical fractionation and immunofluorescence staining¹¹ to localize pp110^{RB} . We used LAN-1 (neuroblastoma) cells for biochemical fractionation because they contain more *RB* protein than any other cell line examined so far. Cells were labelled with ^{35}S -methionine for 3 h and separated by centrifugation into nucleus, cytoplasm and membrane fractions. Equivalent volumes of each sample were immunoprecipitated and analysed by SDS-PAGE. As shown in Fig. *a* most (about 85%) of the ^{35}S -labelled *RB* protein was in the nuclear fraction whereas a minor portion (<10%) was associated with membrane. There was no detectable *RB* protein in the cytoplasmic fraction, or secreted into the medium (data not shown). The significance of *RB* protein associated with membrane is not clear, though it may be due to minor nuclear contamination or to newly synthesized *RB* protein trapped in the membranous apparatus. To substantiate further that the *RB* protein is localized in nucleus, we examined osteosarcoma cell line U2OS, which has optimal cell morphology for immunohistochemical staining. Anti-*RB* IgG or preimmune IgG were applied to cell preparations followed by rhodamine-conjugated goat anti-rabbit IgG (Fig. *b*). Fluorescence was present mainly in the nucleus; the preimmune control was negative.

Lastly, we examined pp110^{RB} for evidence of DNA binding activity. LAN-1 cells were labelled with ^{32}P -phosphoric acid,

and cellular lysates were passed through single- or double-stranded calf thymus DNA/cellulose columns¹². Columns were then eluted with NaCl stepwise from 0.05 M to 1 M. Each fraction was immunoprecipitated with anti-RB IgG and analysed by SDS-PAGE. As shown in Fig. 5, pp110^{RB} was maximally eluted from the single-stranded DNA column at NaCl concentrations of 0.5–1.0 M and from the double-stranded DNA column at 0.3–0.5 M. About 20–40% of total pp110^{RB} reproducibly bound to either column, with the rest of the protein found in the flow-through fraction. These observations have at least two possible explanations: either pp110^{RB} has variable DNA affinity, depending on modification by, or interactions with, other factors; or pp110^{RB} binds to a limited number of DNA sites that are easily saturated. Because the pp110^{RB} used in this assay was not biochemically purified, it is not known whether DNA binding activity is intrinsic to pp110^{RB} or is due to its association with other DNA-binding proteins. In either case, these preliminary data and the nuclear localization results are supportive of a gene-regulatory function for the RB protein. DNA-binding is a readily assayed activity of this protein and may be the most important index of its function.

Several proto-oncogenes such as *c-myc*, *N-myc*, *c-myb* and *c-fos* are nuclear phosphoproteins with DNA binding activity^{12,13}. Oncogenic activation of these proto-oncogenes occurs by deregulation of gene expression or by structural modification, and presence of the gene product is essential for oncogenicity. In contrast, the absence of pp110^{RB} appears to be oncogenic. If pp110^{RB} is important in regulating other genes, deregulation of these genes by loss of pp110^{RB} may mediate

oncogenicity. It will be crucial to identify the target genes presumably regulated by pp110^{RB} as well as other factors which may modulate activity of this protein.

We thank G. Walter, J. Lipsick and A. Hsueh for their suggestions and comments; J. Lipsick for pATH vectors; R. Seeger for neuroblastoma cell lines LAN-1 and LAN-5 and F. Chisari for Alexander hepatoma cell line. This work was supported by a grant from the National Eye Institute (NEI) to W.H.L., a fellowship from the Cancer Research Coordination Committee, University of California, to E.L., and grants from the Elizabeth C. King trust to T.W.S. J.Y.S. is a predoctoral fellow of the Department of Pathology, UCSD. R.E.B. is a recipient of a Physician Scientist Award from the NEI, L.A.D. is the T.D. Duane professor of Ophthalmology, Thomas Jefferson University.

Received 28 July; accepted 28th August 1987.

1. Friend, S. *et al.* *Nature* **323**, 643–645 (1986).
2. Lee, W. H. *et al.* *Science* **235**, 1394–1399 (1987).
3. Spindler, K., Rosser, D. & Berk, J. *Virology* **49**, 132–141 (1984).
4. Boyle, W. J., Lipsick, J. S. & Baluda, M. A. *Proc. natn. Acad. Sci. U.S.A.* **83**, 4685–4689 (1986).
5. Lee, W. H., Wheatley, W., Benedict, W., Huang, C. M. & Lee, E. Y. H. *Proc. natn. Acad. Sci. U.S.A.* **83**, 6790–6794 (1986).
6. Seeger, R. C. *et al.* *New Engl. J. Med.* **313**, 1111–1116 (1985).
7. McFall, R. C., Sery, T. W. & Makedon, M. *Cancer Res.* **37**, 1003–1010 (1977).
8. Tai, T. *et al.* *J. biol. Chem.* **250**, 8569–8575 (1975).
9. Comings, D. E. *Proc. natn. Acad. Sci. U.S.A.* **70**, 3324–3328 (1973).
10. Lee, W. H., Murphree, A. L. & Benedict, B. W. *Nature* **309**, 458–460 (1984).
11. Radke, K. *et al.* *J. Cell Biol.* **97**, 1601–1611 (1983).
12. Ramsay, G., Stanton, L., Schwab, M. & Bishop, J. M. *Molec. Cell Biol.* **6**, 4450–4457 (1986).
13. Weiss, R., Teich, N., Varmus, H. & Coffin, J. *RNA Tumor Viruses* (Cold Spring Harbor Laboratory, New York, (1985).
14. Fung, Y. K. T. *et al.* *Science* **236**, 1657–1661 (1987).

Chromosomal localization of the human rhabdomyosarcoma locus by mitotic recombination mapping

H. J. Scrabble*, D. P. Witte†, B. C. Lampkin‡ & W. K. Cavenee*

* Ludwig Institute for Cancer Research, Royal Victoria Hospital, 687 Pine Avenue West, Montréal, Québec H3A 1A1, Canada

† Division of Pathology and ‡ Division of Hematology/Oncology, Children's Hospital Medical Center, Elland and Bethesda Avenues, Cincinnati, Ohio 45229, USA

A genetic description of the human genome requires maps of three types. The first¹ shows the frequency of chromosomal interchange during meiosis, relying on many equally spaced markers², and is limited to interchanges that do not unmask defects lethal to the conceptus, whose every cell will contain such abnormalities. The second is the physical description of genomic regions defined by karyotypic rearrangements, DNA segments, genes, or their products³. A third description of somatic chromosomal interchanges at mitosis is also required. Because mitotic exchanges occur in a single postembryonic somatic progenitor cell, lethal effects on the organism are reduced. These events have been important in genetic mapping in *Drosophila melanogaster*⁴ and fungi⁵, but they have rarely been detected in mammals^{6–8}. Here we report a significant frequency of mitotic recombination in human tumours and the first application of this information in localizing their predisposing locus.

A model has been proposed⁶ in which predisposing recessive mutations are revealed by chromosomal mechanisms and result in homozygous defects at the tumour locus⁹. These mechanisms include mitotic nondisjunction and loss (with or without reduplication of the remaining homologue), localized gene conversion, point mutation, or small deletion, and mitotic recombination⁶. The existence of mitotic recombination in initiated progenitor

cells giving rise to an observable phenotype (neoplastic growth) suggested that map positions for such loci might be established, even for tumours with no known cytogenetic aberrations, by delineating the smallest overlapping region of somatic homozygosity shared among tumours of similar type.

We have applied this rationale to a genetic description of rhabdomyosarcoma. The presence of rhabdomyoblasts in a paediatric tumour, combined with age and site characteristics, is usually sufficient to diagnose rhabdomyosarcoma¹⁰. No consistent cytogenetic abnormality is apparent in normal or tumour

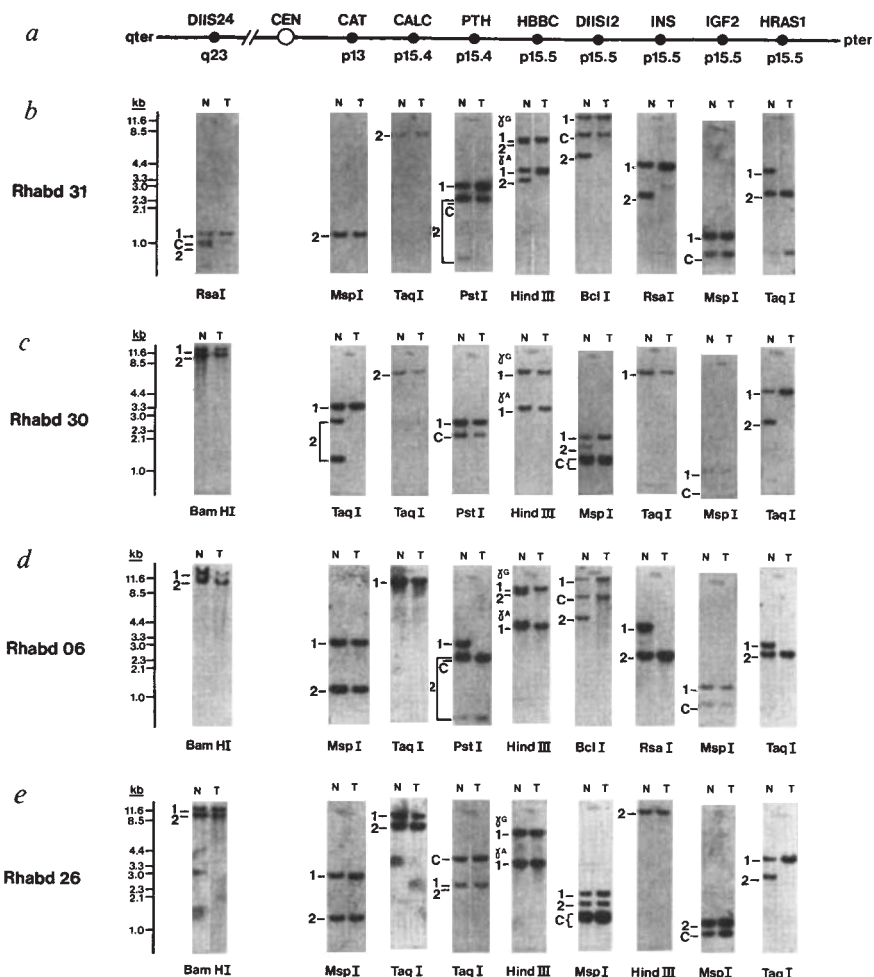
Table 1 Allelic copy number in normal (N) and tumour (T) tissues from rhabdomyosarcoma patients

Patient	Locus	11p/11q ratio		Ratio T/N
		N	T	
Rhabd 31	CAT	1.23	1.33	1.08
	IGF2	0.49	0.53	1.08
	HRAS1	0.56	0.66	1.18
Rhabd 30	CAT	1.67	1.56	0.93
	IGF2	0.80	0.80	1.00
	HRAS1	0.67	0.64	0.96
Rhabd 06	CAT	1.00	1.01	1.00
	IGF2	0.64	0.63	0.98
	HRAS1	1.04	0.96	0.92
Rhabd 26	CAT	1.62	1.64	1.02
	IGF2	2.26	2.20	0.97
	HRAS1	2.76	2.86	1.04

DNA samples from paired normal and tumour tissues from each patient were digested with either *Pvu*II (Rhabd 31, Rhabd 30) or *Pst*I (Rhabd 06, Rhabd 26), electrophoresed, and transferred as described in the legend to Fig. 1. Southern blots were hybridized simultaneously with labelled chromosome 11 short arm^{14,20,21} and long arm¹³ probes. The resultant autoradiograms were analysed by scanning densitometry with an LKB Ultrosan XL Laser Densitometer with internal integrator. The 11p/11q ratios were determined by dividing the area under the curve at the short-arm locus by the area under the curve at the long-arm locus. The T/N ratio was calculated by dividing the 11p/11q ratio in the tumour by that in the normal DNA.

Fig. 1 Alleles present at loci on chromosome 11 in normal and tumour tissues from patients with rhabdomyosarcoma.

Methods. DNA was extracted from normal tissue (peripheral leukocytes; designated 'N') and from primary tumour and/or low passage xenografts (designated 'T') from 9 patients diagnosed (and confirmed upon pathological review by the Intergroup Rhabdomyosarcoma Study) with rhabdomyosarcoma. These samples were digested to completion with restriction endonucleases (Pharmacia), fragments separated by agarose gel electrophoresis, transferred to Nytran membranes (Schleicher and Schuell), hybridized, washed and exposed to X-ray film as described¹¹. Recombinant DNA probes were radiolabelled by nick-translation in the presence of [³²P]dCTP (Amersham, 3,000 Ci mmol⁻¹). Probe pE46-TGH2 is homologous to a locus mapped to 11q22-q23 and reveals alleles of 12 and 9 kilobases (kb) with *Bam*HI (ref. 13), and of 1.2 and 0.9 kb with *Rsa*I. The catalase probe in 11p13 reveals alleles of 3.2 and 2.2+1.1 kb with *Taq*I (ref. 14), and of 2.6 and 1.0 kb with *Msp*I (our results). Probe pH9C is homologous to the *CALC1* locus with alleles of 8.0 and 6.5 kb in *Taq*I-digested DNA (ref. 15). Probe p20.36 reveals alleles at the parathyroid hormone locus in band p15.4 with sizes of 2.6 and 1.9+0.7 kb in *Pst*I digested DNA, and of 1.5 and 2.4 kb using *Taq*I (ref. 16). The probe JW151 is homologous to the beta-globin gene cluster and reveals alleles of 8.0 and 7.2 kb at the gamma-G portion and of 3.5 and 2.7 kb at the gamma-A portion¹⁷ of the locus. pADJ-762 is an arbitrary DNA segment with alleles of 2.3 and 1.8 kb in *Msp*I-digested DNA, and of 11.6 and 4.2 kb in *Bcl*I-digested DNA (ref. 18). The probe pHins-310, homologous to the insulin gene¹⁹, and probe pTBB2 homologous to the *HRAS1* gene²¹, both exhibit alleles of various sizes due to varying numbers of repeated units in the flanking sequences, which are revealed in *Taq*I, *Rsa*I and *Hind*III restriction enzyme digests. The complementary DNA probe to the *IGF2* locus reveals alleles of 0.9 and 0.7 kb with *Msp*I-digested DNA (ref. 20). Alleles at each locus are designated 1 or 2; 1 is the longer allele. A common, invariant band is designated by 'C'. The marker ladder to the left of the Southern data is derived from standards run with each gel.



cells from patients with this disease although limited analyses¹¹ suggested the involvement of a locus grossly mapped to human chromosome 11. Figure 1 shows results obtained when DNA samples from normal and tumour tissues from four of nine rhabdomyosarcoma patients (designated Rhabd 05-07, 10, 25, 26, 30, 31 and 33) were digested with restriction endonucleases, and the resulting fragments separated by electrophoresis through agarose gels and transferred to nylon membranes^{6,9,11,12}. Each sample pair was independently hybridized to nine recombinant DNA probes for loci mapped to 11q22-q23 (*DI1S24*, ref. 13), 11p13 (*CAT*, ref. 14), 11p15.4 (*CALC*, ref. 15; *PTH*, ref. 16) and 11p15.5 (*HBBC*, ref. 17; *DI1S12*, ref. 18; *INS*, ref. 19; *IGF2*, ref. 20; and *HRAS1*, ref. 21). Each of these probes can detect restriction site variation at its cognate locus and can thus distinguish between a maternally or paternally derived chromosome-11 homologue at specific positions.

Constitutional DNA from patient Rhabd 31 was heterozygous at the *DI1S24*, *PTH*, *HBBC*, *DI1S12*, *INS* and *HRAS1* loci; one allele at each of these loci was not detected in his tumour DNA (Fig. 1b). No karyotypic analysis of this tumour was available, but a densitometric comparison of allelic hybridization intensity between the two samples at each locus (see Table 1) showed that these tumour-specific losses of heterozygosity at loci along the length of chromosome 11 were effected by the loss of one entire homologue and the subsequent duplication of the remaining mutant homologue. This mechanism is

outlined in Fig. 2a. Similar results were obtained with four other cases (not shown): Rhabd 05 (losses at *CAT*, *IGF2* and *HRAS1*), Rhabd 25 (losses at *CAT*, *CALC*, *PTH*, *HBBC* and *DI1S12*), Rhabd 07 (losses at *DI1S24*, *CALC*, *DI1S12* and *IGF2*), and Rhabd 33 (losses at *HBBC* and *DI1S12*).

Constitutional DNA from patient Rhabd 30 was heterozygous at the *DI1S24*, *CAT*, *DI1S12* and *HRAS1* loci; his corresponding tumour tissue had lost heterozygosity at each locus on the short arm of chromosome 11 (*CAT*, *DI1S12*, and *HRAS1*) but retained constitutional heterozygosity at the long arm locus, *DI1S24* (Fig. 1c). Karyotypic analysis of the tumour cells showed no structural abnormalities of the chromosomes 11, and densitometric analysis (Table 1) indicated ratios of hybridization intensity at the long and short arm loci equivalent to those in normal tissue. It is therefore likely that the allele losses shown in Fig. 1c arose after a mitotic interchange between the progenitor chromosomes 11 in the region between the *DI1S24* and *CAT* loci which left the long arm region 11q22-11qter intact but caused the loss of genetic heterozygosity at loci on the short arm (Fig. 2c). Although the inadequate number of markers available between the *DI1S24* and *CAT* loci prevented the precise assignment of the point of mitotic recombination, all loci from 11p13-11pter were uniparental in origin. Similarly, constitutional DNA from patient Rhabd 06 was heterozygous at the *DI1S24*, *CAT*, *PTH*, *HBBC*, *DI1S12*, *INS* and *HRAS1* loci, and his corresponding tumour tissue had lost constitutional

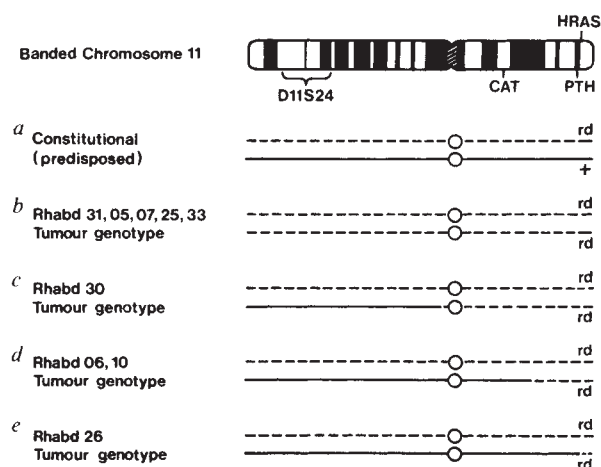


Fig. 2 Schematic model of mitotic recombination mapping of the rhabdomyosarcoma (*Rd*) locus. Chromosome figures are drawn to scale. +, Wild-type allele, *rd*, recessive mutant allele, at the putative rhabdomyosarcoma locus. *a*, Presumed genotype at the *Rd* locus in a cell predisposed to rhabdomyosarcoma either by an inherited (germinal), or somatic mutation at *Rd*. *b*, Representation of chromosome 11 complement in the tumours from patients 31, 05, 07, 25, 33. Loss of wild-type homologue followed by duplication of the mutant, or recombination distal to the outermost marker loci on either arm followed by aberrant segregation, both could result in a cell with this genotype. *c*, Representation of Rhabd 30 genotype. Recombination proximal to *D11S24* followed by aberrant segregation, or proximal to *CAT* followed by normal segregation, results in homozygosity for the entire short arm. *d*, Representation of tumour genotypes for chromosomes 11 in Rhabd 06 and Rhabd 10; recombination occurred between *CAT* (p13) and *PTH* (p15.4). *e*, Recombination within the BAIH linkage group results in reduction to homozygosity for only the most distal region of 11p in Rhabd 26.

heterozygosity at the *PTH*, *HBBC*, *D11S12*, *INS* and *HRAS1* loci but retained both allelic forms of the *D11S24* and *CAT* loci (Fig. 1*d*). A comparison of allelic dosage in his normal and tumour tissue indicated equivalence (Table 1), making it unlikely that the observed allele losses arose by deletion of the distal portion of the short arm of one chromosome 11 homologue. These data are consistent with the model in Fig. 2*d* showing mitotic recombination in the progenitor cell of the tumour at a chromosomal location between the *CAT* (11p13) and *PTH* (11p15.4) loci. Furthermore the Rhabd 10 case showed evidence for mitotic recombination in this same region in the absence of any karyotypically detectable structural abnormality of chromosome 11 (data not shown). The constitutional genotype of patient Rhabd 26 was heterozygous at the *D11S24*, *CAT*, *CALC*, *D11S12* and *HRAS1* loci, whereas her corresponding tumour DNA lost constitutional heterozygosity at the

HRAS1 locus but retained both allelic forms at each of the other loci (Fig. 1*e*). Cytogenetic analysis showed no structural abnormalities of the chromosomes 11 in the tumour and a comparison of allelic dosage in normal and tumour tissue indicated equivalence (Table 1). These data are consistent with the model in Fig. 2*e* showing mitotic recombination in the tumour progenitor cell at a chromosomal location between the two tightly linked loci, *D11S12* and *HRAS1*.

We have thus compared constitutional and tumour genotypes for loci on chromosome 11 in nine cases of histopathologically confirmed rhabdomyosarcoma. Loss of constitutional heterozygosity was observed in each case at loci on chromosome 11p, which supports preliminary suggestions¹¹ that this chromosome contains a locus in which recessive mutations can predispose to the disease. These studies have used mitotic recombination mapping to localize this locus between 11p15.5 and 11pter, the smallest region of overlapping somatic homozygosity in the tumours. Our finding that all tumours lost heterozygosity, and that this resulted in ~50% of them from mitotic recombination, implies that markers for meiotic linkage map construction²² can also be used to localize genes with mutant forms that predispose carriers to strictly somatic phenotypes, and that mapping by mitotic recombination can provide complementary information.

The present data imply that the clinical association of the Beckwith-Wiedemann syndrome (BWS) with increased risk for rhabdomyosarcoma could result from the pleiotropic expression of a single mutant gene in 11p15.5-11pter, for BWS is associated with cytogenetic aberrations of band 11p15^{23,24}. Also, the oncogenetic relation of Wilms' tumour and rhabdomyosarcoma might not result from the pleiotropic expression of a single recessive mutation in 11p13, but rather from mutations in separate loci for Wilms' tumour/aniridia/genito-urinary abnormalities/mental retardation (WAGR) in 11p13²⁵⁻²⁷ and rhabdomyosarcoma in 11p15.5-11pter. This association could be explained by homozygosity of an extended region of the short arm of chromosome 11 which encompasses both loci, or by a rearrangement of a tumour chromosome that brings the loci into juxtaposition and subjects them to coordinate control. It is possible that Wilms' tumour is actually a disparate group of paediatric tumours of the kidney, a subset of which is associated with the WAGR syndrome on 11p13, and a different subset with BWS and rhabdomyosarcoma in 11p15. The present data suggest that approaches similar to those reported here could provide information to distinguish among these possibilities.

We thank Drs H. Kazazian, M. Wigler, R. White, G. Bell, D. Housman, R. Gravel, J. Höppener, H. Mayer and M. Jansen for recombinant DNA probes; Drs S. Soukup and S. Wang-Wuu for karyotypic evaluations; Drs W. Newton and K. Shimada for pathological review; Drs A. Koufos and M. Hansen for discussion; M. Brown, T. Ledford and J. Hansen for technical assistance and E. Lacaille, S. Caluori and L. Sapienza for manuscript preparation. These studies were supported in part by a grant from the National Cancer Institute, NIH.

Received 27 April; accepted 13 August 1987.

- White, R. L. *et al. Nature* **313**, 101-105 (1985).
- Botstein, D., White, R. L., Skolnick, M. & Davis, R. W. *Am. J. hum. Genet.* **32**, 314-331 (1980).
- Shows, T. B., Sakaguchi, A. Y. & Naylor, S. L. *Adv. hum. Genet.* **12**, 341-452 (1982).
- Postlethwait, J. H. in *Genetics and Biology of Drosophila*, Vol. 2c (eds Ashburner, M. & Novitski, E.) 359-343 (Academic, New York, 1978).
- Pontecorvo, G. & Kafer, E. *Adv. Genet.* **9**, 71-104 (1958).
- Cavenee, W. K. *et al. Nature* **305**, 779-784 (1983).
- Wasmuth, J. J. & Hall, L. V. *Cell* **36**, 697-707 (1984).
- Raizis, A. M., Becroft, D. M., Shaw, R. L. & Reeve, A. E. *Hum. Genet.* **70**, 344-346 (1985).
- Cavenee, W. K. *et al. Science* **228**, 501-503 (1985).
- Enzinger, F. & Weiss, S., eds in *Soft Tissue Tumours* 338-378 (Mosby, St Louis, 1983).
- Koufos, A. *et al. Nature* **316**, 330-334 (1985).
- Hansen, M. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 6216-6220 (1985).
- Glaser, T. *et al. Cytogenet. Cell Genet.* **40**, 643 (1985).

- Quan, F., Korneluk, R. G. & Gravel, R. A. *Am. J. hum. Genet.* **37**, A171 (1985).
- Höppener, J. *et al. Hum. Genet.* **66**, 309-312 (1984).
- Schmidtke, J. *et al. Hum. Genet.* **67**, 428-431 (1984).
- Antonarakis, S. E., Boehm, C. D., Giardina, P. J. & Kazazian, H. H. *Proc. natn. Acad. Sci. U.S.A.* **79**, 137-141 (1982).
- Barker, D., Holm, T. & White, R. L. *Am. J. hum. Genet.* **36**, 1159-1171 (1984).
- Bell, G. I., Karam, J. H. & Rutter, W. J. *Proc. natn. Acad. Sci. U.S.A.* **78**, 5759-5763 (1981).
- Jansen, M. *et al. FEBS Lett.* **179**, 243-246 (1985).
- Goldfarb, M., Shimizu, K., Perucho, M. & Wigler, M. *Nature* **296**, 404-409 (1982).
- Willard, H. F., Skolnick, M. H., Pearson, P. L. & Mandel, J. L. *Cytogenet. Cell Genet.* **40**, 360-489 (1985).
- Waziri, M., Patil, S. R., Hanson, J. W. & Bartley, I. A. *J. Pediatr.* **102**, 873-876 (1983).
- Turleau, C. *et al. Hum. Genet.* **67**, 219-221 (1984).
- Knudson, A. G. & Strong, L. C. *J. natn. Cancer Inst.* **48**, 313-324 (1972).
- Riccardi, V. M. *et al. Pediatrics* **61**, 604-610 (1978).
- Francke, U. & Riccardi, V. M. *Cytogenet. Cell Genet.* **24**, 185-192 (1979).

Multiple *cis*- and *trans*-acting elements mediate the transcriptional response to phorbol esters

Robert Chiu, Masayoshi Imagawa, Richard J. Imbra*, John R. Bockoven & Michael Karin

Department of Pharmacology, School of Medicine and Center for Molecular Genetics, University of California, San Diego, La Jolla, California 92093, USA

Protein kinase C is important in the transduction of signals generated at the plasma membrane. The physiological activators of protein kinase C are diacylglycerols¹, and the tumour-promoting phorbol esters, such as 12-*O*-tetradecanoyl-phorbol-14 acetate (TPA), constitute another group of specific activators^{1,2}. Many cellular substrates for phosphorylation by protein kinase C have been described^{1,3}, but proteins that directly control transcription in response to protein kinase C activation are yet to be identified. TPA treatment leads to induction of various proto-oncogenes⁴⁻⁷, growth factor genes^{7,8}, and genes encoding secreted proteases⁹. In addition, TPA increases the activity of viral enhancer elements¹⁰⁻¹³. To identify *trans*-acting factors that mediate the transcriptional response to TPA we chose the simian virus 40 (SV40) enhancer¹⁴ as a model, because it is known to be composed of several discrete *cis*-acting elements¹⁵⁻¹⁸ which are recognized by multiple *trans*-acting factors¹⁹⁻²¹. We report here that the SV40 enhancer contains at least four different TPA responsive elements whose activity is dependent on cell-type. The induction response is likely to involve at least two distinct post-translational steps which modulate the activity of the proteins that recognize these elements.

Analysis of several TPA inducible genes revealed a consensus sequence, 5'-TGAGTCAG-3', which functions as a TPA responsive element (TRE), conferring inducibility to heterologous promoters²². This sequence is recognized by transcription factor AP-1 (refs 22, 23). To determine the function of AP-1 in activation of the SV40 enhancer, we compared the activity of the wild-type enhancer in plasmid pAO to that of the mutant enhancer in pA32, which contains three point mutations in the SV40 P motif (ref. 15; Fig. 1a). DNase I footprinting shows that these mutations severely reduce protection of this site by AP-1 (Fig. 1b), which is present in extracts of control or TPA-treated HepG2 cells. These mutations do not affect either the basal expression of the SV40- β -globin gene or its induction by TPA (Fig. 1c). From these experiments it seems that AP-1 binding is not required for induction of SV40 enhancer activity by TPA in transiently transfected HepG2 cells.

To determine whether other elements of the enhancer are involved in activation by TPA, we examined other mutants containing double point mutations in the A, B or C elements¹⁶. All these mutants showed a dramatic decrease in enhancer function and were refractory to TPA induction (data not shown), indicating that in HepG2, as in HeLa cells^{15,16}, the A, B and C elements are more important for enhancer function than the P motif. In addition, one or all of these elements may function as TREs.

Ondek *et al.*¹⁷ recently constructed synthetic enhancers composed of multiple tandem copies of the A, B or C elements. Analysis of constructs containing these elements downstream of the β -globin gene showed that each wild-type element can independently enhance transcription and that mutant elements are inactive. We used these constructs to determine which of the elements exhibits TRE activity. In addition, we examined the activity of a multimer of the P motif (5 $\times$ P: see Fig. 2b legend

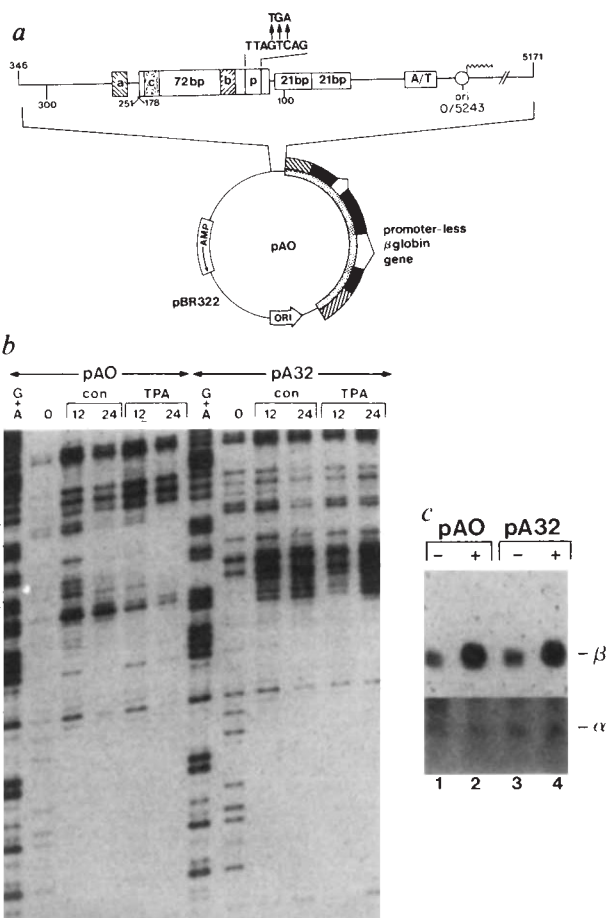


Fig. 1 The role of AP-1 binding in the response of the SV40 enhancer to TPA. *a*, Structure of the constructs used to test the role of AP-1 binding in enhancer activation. pAO is the basic construct containing the SV40 early promoter and enhancer fused to a promoter-less rabbit β -globin gene. This construct is described in detail by Zenke *et al.*¹⁵. pA32 is a derivative of pAO in which the triplet GTC at positions 189-187 within the AP-1 binding site of the SV40 enhancer is replaced by the triplet TGA (ref. 15). *b*, Effect of the A32 mutation on AP-1 binding. To examine the binding of AP-1 to pAO and pA32, the two plasmids were digested with *Nco*I (position 37 of SV40 DNA), labelled by filling-in and subjected to secondary digest with *Asp*718 (position 294). The *Nco*I-*Asp*718 fragments of the two plasmids were isolated and 10,000 c.p.m. (~ 2 ng) of each were incubated with 0, 12 or 24 μ g of protein from partially purified HepG2 whole cell extracts prepared from control or TPA-treated (100 ng ml⁻¹; 1 h) cells, and subjected to DNase I footprinting as described²². The location of the various binding sites is indicated on the side panel. *c*, Expression of pAO and pA32. Total cellular RNA isolated from HepG2 cells transfected with 5 μ g of either plasmid together with 5 μ g of α -globin, used as an internal control, was analysed by Northern blot hybridization (40 μ g per lane) using rabbit β -globin and human α -globin specific probes as described¹⁰. The treated cells (+) were incubated with TPA (100 ng ml⁻¹) for 4 h before RNA extraction. The correct sized transcripts derived from the SV40/ β globin and α -globin genes are indicated as β and α (correct initiation of these transcripts was verified by S₁ nuclease protection; data not shown).

for enhancer notation), which was also inserted downstream to the β -globin gene (Fig. 2a). These plasmids were transfected into HepG2 cells and the amount of β -globin messenger RNA in control and TPA-treated cells were monitored by RNase protection. Each of the elements had TRE activity (Fig. 2b) but the relative activities of the elements were different. After induction, 5 $\times$ P is as active as the complete SV40 enhancer (1 $\times$ 72), 6 $\times$ C is much more active and 6 $\times$ B is the least active. A hexameric repeat of the A element also showed TRE activity,

* Present address: New York University Medical Center, Department of Environmental Medicine, New York, New York 10016, USA.

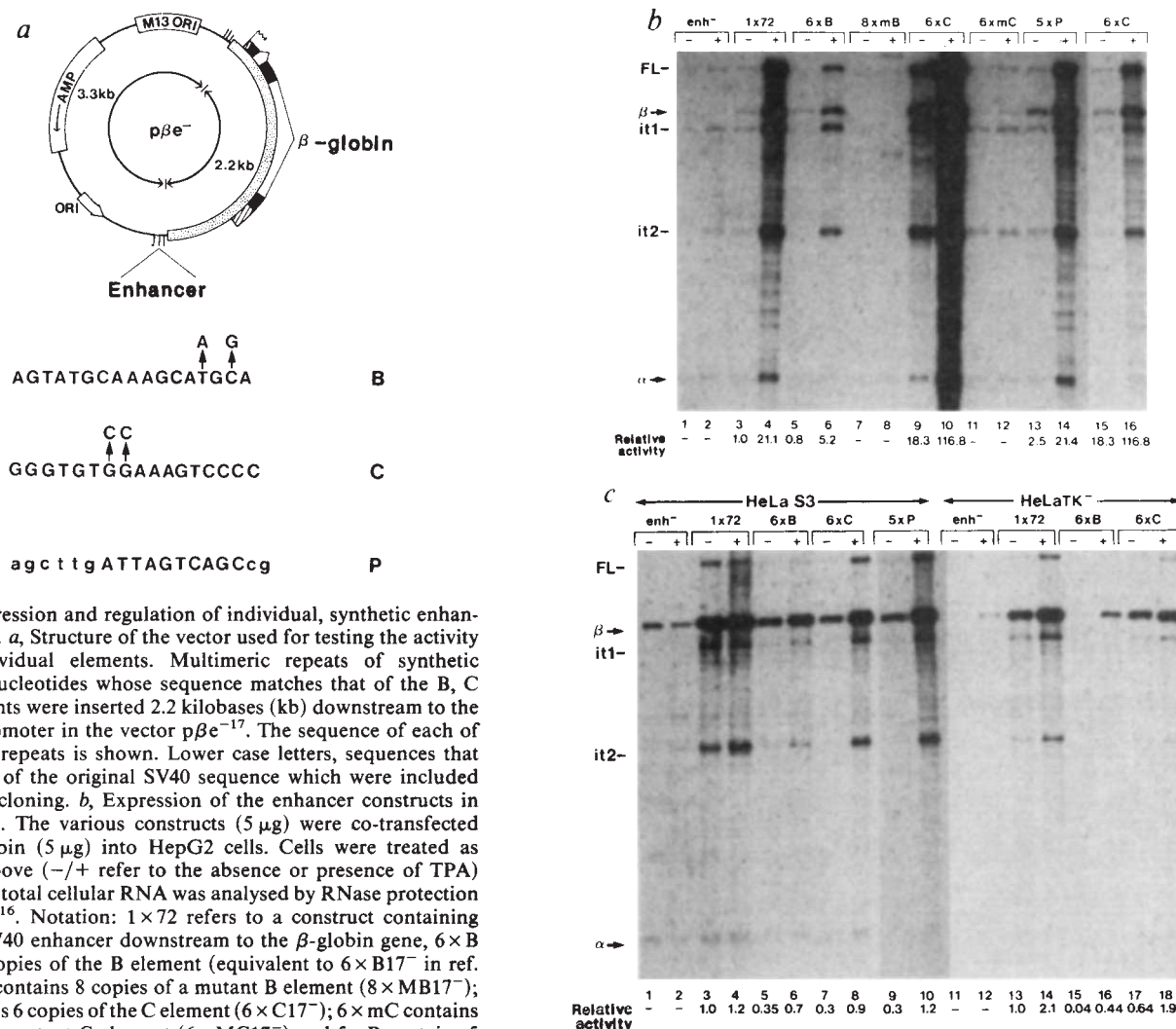


Fig. 2 Expression and regulation of individual synthetic enhancer elements. **a**, Structure of the vector used for testing the activity of the individual elements. Multimeric repeats of synthetic oligodeoxynucleotides whose sequence matches that of the B, C and P elements were inserted 2.2 kilobases (kb) downstream to the β -globin promoter in the vector p β e⁻. The sequence of each of the inserted repeats is shown. Lower case letters, sequences that are not part of the original SV40 sequence which were included to facilitate cloning. **b**, Expression of the enhancer constructs in HepG2 cells. The various constructs (5 μ g) were co-transfected with p α -globin (5 μ g) into HepG2 cells. Cells were treated as described above (-/+ refer to the absence or presence of TPA) and 20 μ g of total cellular RNA was analysed by RNase protection as described¹⁶. Notation: 1x72 refers to a construct containing the intact SV40 enhancer downstream to the β -globin gene, 6xB contains 6 copies of the B element (equivalent to 6xMB17⁻ in ref. 17); 8xmB contains 8 copies of a mutant B element (8xMB17⁻); 6xC contains 6 copies of the C element (6xC17⁻); 6xmC contains 6 copies of a mutant C element (6xMC17⁻) and 5xP contains 5 copies of the P motif. Lanes 15 and 16 show a shorter exposure of lanes 9 and 10. **c**, The same constructs were transfected into HeLa S3 and HeLa TK⁻ cells, and total cellular RNA (10 μ g) from TPA treated (+) and untreated (-) cells was analysed for β - and α -globin expression as described above. The transfection protocol for HeLa S3 and HeLa TK⁻ cells was slightly different from the one used for HepG2 cells. Instead of a 4 h incubation with the calcium-phosphate-DNA coprecipitate and a glycerol shock¹⁰, the cells were incubated with the coprecipitate for 12–14 h, after which they were washed and incubated in medium containing 0.5% fetal calf serum for 24 h. TPA was added to 100 ng ml⁻¹ for a further 4 h and the cells collected. α and β in panels B and C refer to correctly initiated α - and β -globin transcripts, while it1 and it2 refer to incorrect transcripts that probably result from improper splicing of aberrant transcripts¹⁶. FL refers to a transcript initiated upstream to the correct β -globin start site which leads to protection of the full length β -globin region of the probe. The relative activities of the various constructs shown at the bottoms of panels B and C were determined by densitometric analysis of the autoradiograms shown here and from another independent transfection experiment, taking into consideration only the correct transcripts, and averaging the two.

comparable to that of 6xB (data not shown). Although all the elements are TPA responsive, the effect is specific because point mutants of the B and C elements (8xmB and 6xmC) and the co-transfected α -globin gene are unresponsive. (The apparent increase in the α -globin signal in lanes 4, 9, 10 and 14 is due to the high background from degradation products of β -globin mRNA.)

The same constructs were used to investigate the lack of response of the SV40 enhancer to TPA in HeLa S3 cells¹⁰. The intact enhancer is constitutively active in HeLa S3 cells, whereas 5xP exhibits TRE activity (Fig. 2c). Unlike the intact enhancer, 6xC and 6xB are TPA inducible, but not to the same extent as they are in HepG2 cells. HeLa S3 cells multiply more rapidly than HepG2 cells and originate from different carcinomas. The activity of these elements could be influenced by growth control, so we examined their inducibility in a slower growing variant of HeLa S3 cells, HeLa TK⁻ (ref. 24). As shown in Fig. 2c, the basal and induced activities of 5xP and 6xC in HeLa TK⁻ are similar to their activity in HeLa S3 (data not shown for 5xP in

HeLa TK⁻). In contrast, the basal activity of 6xB is very low in HeLa TK⁻ cells but is markedly increased by TPA. Consequently, the intact enhancer is also moderately responsive to TPA in HeLa TK⁻ cells, but is refractory to induction in HeLa S3. In both HeLa lines the behaviour of the A element is similar to that of the C element (data not shown). Because the thymidine kinase (TK) gene is not required for growth under normal conditions, we do not believe that the TK⁻ genotype is directly responsible for these differences. It is not understood why HeLa TK⁻ cells grow more slowly than HeLa S3, and have a greater dependence on serum (data not shown).

These *cis* elements may function as recognition sites for *trans*-acting factors that transmit the signal elicited by TPA to the transcriptional machinery. To investigate the effect of TPA on the factors that recognize these elements, protein extracts were prepared from control and TPA-treated cells and the extent of DNA-binding activity was monitored by footprint titration experiments (Fig. 3). In agreement with earlier observations²², treatment of HepG2 and HeLa S3 cells with TPA results in

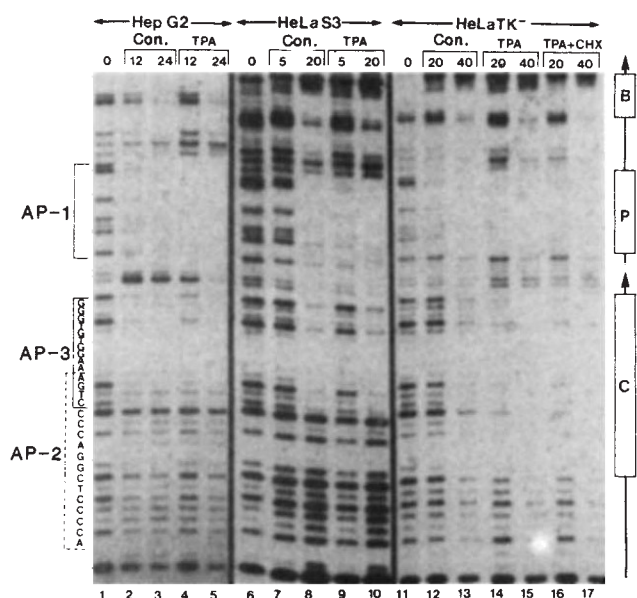


Fig. 3 Effect of TPA treatment on interaction of cellular proteins with the SV40 enhancer. Whole cell extracts prepared as described²² from untreated control (Con.) or treated cells (with TPA for 1 h in the absence or presence of 30 $\mu\text{g ml}^{-1}$ cycloheximide (CHX)) were partially enriched for DNA binding proteins on heparin agarose columns. The cells were grown under the same conditions used for the transfection experiments above. To detect the presence of DNA binding proteins which recognize the SV40 enhancer, the DNaseI footprinting procedure was used²². About 2 ng of an SV40 probe, end-labelled at the *Bam*HI site (+101) by 'filling-in', was incubated with the indicated amounts (in μg) of partially purified extracts and treated with DNaseI. The different binding sites were determined by using affinity-purified proteins (ref. 22; Fig. 4b) and the location of the various elements within the enhancer are indicated. Note that the region shown contains the end of the first repeat (the C box) and the beginning of the second repeat (the P and B boxes). The sequences protected by AP-2 and AP-3 are indicated at left (not to scale).

increased AP-1 binding activity, the B and C element binding activities remaining constant. In HeLa TK⁻ cells TPA leads to increased C-element binding activity, which is independent of *de novo* protein synthesis, as judged by its insensitivity to cycloheximide. The difference in C-element binding activity contrasts with its similar transcriptional response to TPA in HeLa S3 and HeLa TK⁻ cells. These differences were consistently observed in three separate experiments. No protection of the A element was observed using unfractionated extracts.

Because at least two distinct factors bind to different sequence motifs in the C box (Fig. 4b), it was important to determine which of them was responsible for activation of this element. We therefore compared protection of the wild-type enhancer with that of the *dpm6* mutant¹⁶. Consistent with its deleterious effect on the *in vivo* activity of the C element (Fig. 2), the *dpm6* mutation causes a decrease in protection by extracts of control or TPA-treated cells (Fig. 4a). These results suggest that the C element is protected by a factor whose recognition sequence includes the two G-C base pairs affected by the *dpm6* mutation. This binding activity is constitutive in HeLa S3 and inducible in HeLa TK⁻ cells.

This factor was partially purified from HeLa S3 extracts for further characterization. This protein, activator protein 3 (AP-3), shows decreased binding to the *dpm6* mutant (Fig. 4b), as would be predicted for the factor responsible for activation of the C element. AP-3 purified by affinity chromatography gives footprints similar to those from crude HeLa S3 and HepG2 extracts. These are clearly distinct from the footprints AP-2, which also binds to the C element (M.I. *et al.*, in preparation; P. Mitchell *et al.*, in preparation). The *dpm6* mutation has no

effect, however, on binding of AP-2 (M.I. *et al.*, in preparation). Therefore, it appears that AP-3 and not AP-2 is responsible for activation of the C element *in vivo*.

Compared to the region protected by purified AP-3, protection of the C element by the TPA induced HeLa TK⁻ extract extends one to two base pairs further toward the AP-2 site (see Figs 3 and 4a). The mobility of the complex formed by the inducible factor in HeLa TK⁻ cells and a C-element probe was found to be identical to the mobility of the complex formed with the constitutive HeLa S3 factor (data not shown), suggesting that both cells contain the same C-element binding activity. The reason for the slight differences between the footprints is not yet clear, but it is possible that they are caused by a TPA-induced post-translational modification of AP-3 in HeLa TK⁻ cells.

These results indicate that the increased efficacy of the intact SV40 enhancer after TPA treatment is a result of the combined action of the A, B and C elements. These all have TRE activity, whereas the P motif harbouring the first described TRE (ref. 22) contributes very little to the activity of the intact enhancer. The P motif alone is a potent TRE, which suggests that in the intact enhancer its activity could be repressed by a neighbouring element. Inducibility of the P motif does not vary significantly between the three cell lines we examined, but the response of the A, B and C elements to TPA is dependent on cell-type. The basal activities of these elements are also cell-type dependent^{17,18}. This can arise from differences in the relative amounts of activator proteins and other factors that modulate their activity, such as PKC (ref. 25).

Initially, the observation that each of the individual elements examined has TRE activity was surprising and suggested a lack of specificity. But we report here and elsewhere^{10,22} that a number of other promoters are not inducible and therefore *cis* elements, such as the GC, CAAT and TATA boxes, do not have TRE activity. Further support for the specific nature of the response is provided by its cell-type dependence. It might be expected that several distinct elements should be involved in the transcriptional response to phorbol esters because of the range of effects these agents have on animal cells^{1,3}.

In addition to the four factors proposed in this paper, another transacting factor, NF κ B, may also mediate gene induction by TPA and bacterial lipopolysaccharide. In mature B cells NF κ B is expressed constitutively, but in pre-B cells its binding activity is induced by TPA or bacterial lipopolysaccharide^{26,27}. Preliminary results suggest that NF κ B and AP-3 are distinct proteins in spite of their similar binding sites (M.I. and K. Jones, unpublished results). Thus, at least five different DNA-binding proteins are part of the TPA-activated signal transduction network. The multiplicity of elements involved in the TPA response stands in contrast to gene induction by heavy metal ions or glucocorticoid hormones²⁸, for which only a single prototype of responsive element has been found, though the actual sequence may vary^{29,30}.

The biochemical events that modulate the activity of these factors are yet to be determined, but from our present results two basic mechanisms are feasible. One mechanism applies to AP-1 and other factors whose DNA-binding activity is increased by TPA. This increase may be caused by post-translational modifications and is sufficient to account for transcriptional stimulation²². Such a mechanism cannot apply for AP-3, whose activity is inducible in HeLa TK⁻ but not in HeLa S3 or HepG2 cells, suggesting that at least two distinct steps are involved in modulation of AP-3 activity. In the first step the DNA-binding activity of the factor is increased by a post-translational event, which follows from its independence of *de novo* protein synthesis. But increased DNA-binding cannot solely account for induction because in HeLa S3 and in HepG2 cells in which AP-3 binding activity is constitutive, the C element is fully inducible. These and the results obtained with the A and B elements implicate a second step responsible for increasing the transcriptional stimulatory activity of the factors, but not their

Fig. 4 Effect of point mutations on protection of the C element. *a*, Effect of the *dpm6* mutation on protection of the C element by whole cell extracts. The 'wild-type' and *dpm6* probes used in this experiment were labelled at the Asp718 site (+294) by 'filling-in' (this is the opposite strand to the one shown in Fig. 3) and were incubated with 20 µg of partially purified extracts of TPA treated (+) and untreated (-) HeLa S3 and HeLa TK⁻ cells. 0-no extracts were added. The products of the G+A chemical sequencing reaction were used as markers. *b*, Binding of affinity-purified AP-3 and AP-2 to the wild-type (WT) enhancer and to the *dpm6* mutant. The wild-type probe was incubated with 8 µl AP-3 and 8 µl AP-2 preparations; the *dpm6* probe was incubated with 8 (+) and 16(++) µl of the AP-3 preparation. Note that the sequences protected by AP-2 and AP-3 are not to scale.

Methods. AP-2 and AP-3 were purified by sequence specific affinity chromatography³¹ after fractionation of HeLa S3 whole cell extracts on heparin-agarose and Sephacryl S-300 columns, as described for purification of AP-1 (ref. 22). The sequences of the binding sites used for preparing the affinity columns are:

AP2: 5' GATCGAAGTACCGCCGCGGCCCGT
CTTGACTGGCGGCGCGGCCGACTAG

and

AP3: 5' GATCTGTGGAAGTCCCA
ACACCTTTCAGGGTCTAG

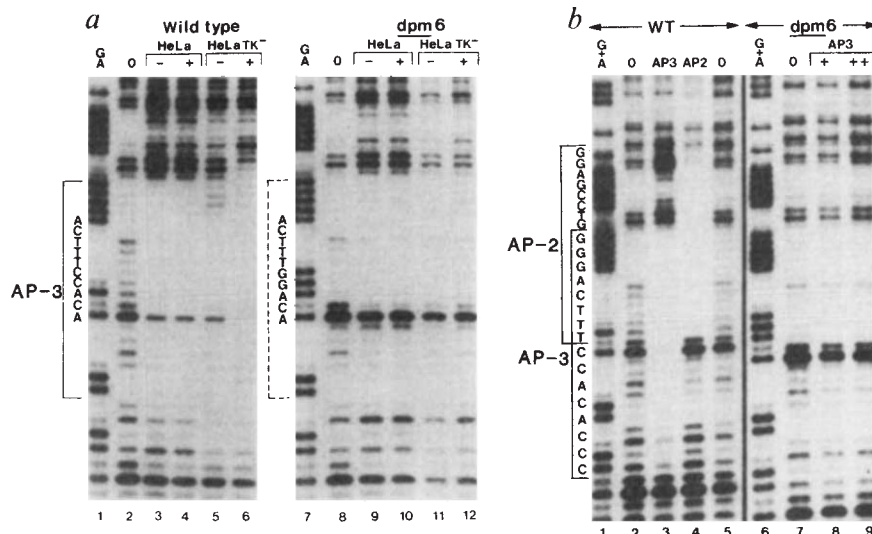
A detailed description of the purification of these proteins will be published elsewhere (M.I. *et al.*, in preparation).

affinity for DNA. Thus, activator proteins can be present in certain cells in a form that, in spite of DNA binding, cannot activate transcription efficiently, and in other cell types they are present in a non-binding, non-active form. Both forms can be activated after exposure to hormones or phorbol esters by post-translational events that may involved protein phosphorylation.

We thank P. Chambon and W. Herr for the mutant SV40 enhancers, B. Ondek and W. Herr for the multimerized A, B and C element constructs and communication of results before publication, K. Jones and T. Hunter for critical reading of this manuscript, E. Satuloff and H. To for technical assistance, J. Meek for artwork and P. Bucher for typing. Research was supported by the National Institute of Environmental Health Sciences, Environmental Protection Agency, Department of Energy and Council for Tobacco Research. M.K. is a Searle Scholar and B.C. is supported by an NIH PHS postdoctoral fellowship.

Received 11 May; accepted 5 August 1987.

- Nishizuka, Y. *Nature* **308**, 693-698 (1984).
- Blumberg, P. M. *CRC crit. Rev. Tox.* **9**, 153-197 (1981).
- Nishizuka, Y. *Science* **233**, 305-312 (1986).
- Greenberg, M. E. & Ziff, E. B. *Nature* **311**, 433-438 (1984).
- Kruijer, W., Cooper, J. A., Hunter, T. & Verma, I. M. *Nature* **312**, 711-716 (1984).
- Kelley, K., Cochran, B. H., Stiles, C. D. & Leder, P. *Cell* **35**, 603-610 (1983).
- Colamonici, R., Trepel, J. B., Vidal, C. A. & Neckers, L. M. *Molec. cell Biol.* **6**, 1847-1850 (1986).
- Fujita, T., Ohno, S., Yasumitsu, H. & Taniguchi, T. *Cell* **41**, 489-496 (1985).
- Angel, P. *et al. Molec. cell Biol.* **7**, 2256-2266 (1987).
- Imbra, R. J. & Karin, M. *Nature* **323**, 555-558 (1986).
- Wasylyk, C., Imil, J. L., Peres-Mutul, J. & Wasylyk, B. *Cell* **48**, 525-534 (1987).
- Nabel, G. & Baltimore, D. *Nature* **326**, 711-714 (1987).
- Dinter, H., Chiu, R., Imagawa, M., Karin, M. & Jones, K. *EMBO J.* (submitted).
- Khoury, G. & Gruss, P. *Cell* **33**, 313-314 (1983).
- Zenke, M. *et al. EMBO J.* **5**, 387-397 (1986).
- Herr, W. & Clarke, G. *Cell* **45**, 461-470 (1986).
- Ondek, B., Shepard, A. & Herr, W. *EMBO J.* **6**, 1017-1025 (1987).
- Schirm, S., Jiricny, J. & Schaffner, W. *Genes Dev.* **1**, 65-74 (1987).
- Wildeman, A. G. *et al. Molec. cell Biol.* **6**, 2098-2105 (1986).
- Bohman, D. *et al. Nature* **325**, 268-272 (1987).
- Lee, W., Haslinger, A., Karin, M. & Tjian, R. *Nature* **325**, 368-372 (1987).
- Angel, P. *et al. Cell* **49**, 729-739 (1987).
- Lee, W., Mitchell, P. & Tjian, R. *Cell* **49**, 741-752 (1987).
- Kit, S., Dubbo, D. R. & Fearson, P. M. *Int. J. Cancer* **1**, 19-30 (1966).
- Carpenter, D., Jackson, T. & Hanley, M. R. *Nature* **325**, 107-108 (1987).
- Sen, R. & Baltimore, D. *Cell* **47**, 921-928 (1986).
- Atchison, M. L. & Perry, R. P. *Cell* **48**, 121-128 (1987).
- Karin, M. *et al. Nature* **308**, 513-519 (1984).
- Karin, M., Haslinger, A., Heguy, A., Dietlin, T. & Cooke, T. *Molec. cell Biol.* **7**, 606-613 (1987).
- Janzen, H.-M. *et al. Cell* **49**, 29-38 (1987).
- Kadonaga, J. T. & Tjian, R. *Proc. natn. Acad. Sci. U.S.A.* **83**, 5889-5893 (1986).



Similarity between cell-cycle genes of budding yeast and fission yeast and the *Notch* gene of *Drosophila*

Linda Breeden* & Kim Nasmyth

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

The *HO* gene of *Saccharomyces cerevisiae* encodes the endonuclease¹ that initiates mating-type switching². To prevent inopportune switching, *HO* transcription is restricted to a specific period in the haploid³⁻⁵ cell cycle, which is just after, and dependent on, the start of the mitotic cell cycle⁶. A repeated promoter element (CACGA4) (refs 7-9) and two *trans*-acting activators (*SWI4* and *SWI6*)⁹ have been identified, which are responsible for the periodic and start-dependent transcription of *HO*. To understand further the link between start and *HO* transcription, the *SWI6* gene has been cloned and sequenced. The *SWI6* protein is similar to the protein in *Schizosaccharomyces pombe* that is encoded by *cdc10* an essential gene specifically required at the start of the cell cycle^{12,13}. The similarity between the *SWI6* and *cdc10* products, and their common involvement with 'start', suggest that they may share a common mechanism for sensing or executing this critical control step in the cell cycle. The *SWI6* and *cdc10* proteins also contain two copies of a repeated motif that occurs at least five times in the cytoplasmic domain of the *Notch* protein of *Drosophila melanogaster*.

The sequence of the *SWI6* gene is shown in Fig. 1. It contains one long open reading frame with no evidence of introns. The product of its translation would be a highly soluble polypeptide 803 amino acids long, with a relative molecular mass (M_r) of 90,561. This predicted polypeptide would have no strongly hydrophobic domains, either internally or at the amino-terminus. The codons used show no bias toward those that predominate in the very abundant proteins of *Saccharomyces cerevisiae*¹⁰.

* Present address: Fred Hutchinson Cancer Research Center, 1124 Columbia Street, Seattle, Washington 98104, USA.



Fig. 1 Cloning and sequencing of the *SWI6* gene. Above, restriction map of the *SWI6* gene, showing the subclones and insertion mutants generated in relation to the *SWI6* open reading frame (ORF). The rightmost column indicates whether or not the DNA fragment complements a *swi6* mutation. Solid bars, inserted DNA; open bars, deleted DNA. Below, nucleotide sequence and derived protein sequence of the *SWI6* gene. Nucleotide sequences are shown extending 2.85 kb from the *HindIII* to *BglII* sites.

Methods. DNA fragments that complemented the *swi6-399* mutation were isolated from gene banks made in a vector based on the 2- μ m plasmid (yEP13) and a *CEN* vector (yCP50). Complementation was inferred by transforming³² these banks into a *swi6-399 ho: lacZ46* strain and screening transformants for restoration of β -galactosidase activity^{8,9}. Restriction mapping and hybridization studies (data not shown) showed that the complementing DNAs were from two different chromosomal locations. To determine which of these complementing DNAs contained the *SWI6* gene, they were subcloned and then deletions within them were made. The deletions were genetically marked by insertion of either the *LEU2* or *TRP1* gene, and were integrated into the yeast genome by homologous recombination to replace the wild-type gene³³. Integration of the deleted genes shown here cause *Swi*⁻ phenotypes that map to the *SWI6* locus. This was determined as follows. A linear DNA fragment carrying the *LEU2* gene cloned into the *ClaI*-*BglII* *SWI6* deletion, was transformed into a diploid strain (*ho: lacZ46, leu2*) and *leu2* transformants were selected. When these diploids were sporulated, they showed a segregation pattern of 2 *Leu*⁺ *Swi*⁻ : 2 *Leu*⁻ *Swi*⁺ in all tetrads analysed. The DNA from a *Leu*⁺ *Swi*⁻ haploid was analysed by Southern hybridizations³⁴ to confirm that a simple replacement of the wild-type locus had taken place. This *Swi*⁻ haploid was then crossed to a *swi6-399, ho: lacZ46, leu2* strain, sporulated, and the tetrads were dissected. Of the 44 tetrads analysed, 43 yielded 4 *Swi*⁻ spores. The exceptional tetrad yielded 2 *Swi*⁺ spores. All four spores from this tetrad were *Leu*⁻, suggesting that the *swi6* deletion had been lost by mitotic recombination. Because no *Swi*⁺ *Leu*⁺ recombinants were found, the insertion must either be very tightly linked to, or allelic with, *swi6-399*. The *SWI6* sequence was obtained by dideoxy sequencing³⁵ using processive deletions generated by *exoIII* and *S1* digestion³⁶. Both strands of the DNA were sequenced in full. The second DNA fragment that complemented the *swi6-399* mutant (not shown) encodes the *SWI4* gene. This was determined as above.

A search¹¹ has been made through all available databases for proteins with sequences related to *SWI6*. The most significant similarity detected was that of the product of the *cdc10* gene¹² from the distantly related fission yeast *Schizosaccharomyces pombe*. The *cdc10* gene is one of two that have been identified that are required to start the cell cycle in *S. pombe*¹³. The other known prerequisite for start (the *cdc2* gene) has been shown to

be both structurally and functionally homologous to the *cdc28* gene of *Sacc. cerevisiae*^{14,15} and more recently, to have a homologue in humans^{16,17}. The finding of a second pair of related, start-specific proteins suggests that start may be a highly conserved control step in the cell cycle.

The positions and extent of the similarity between *SWI6* and *cdc10* are shown in Fig. 2. It is apparent from the DIAGON¹⁸

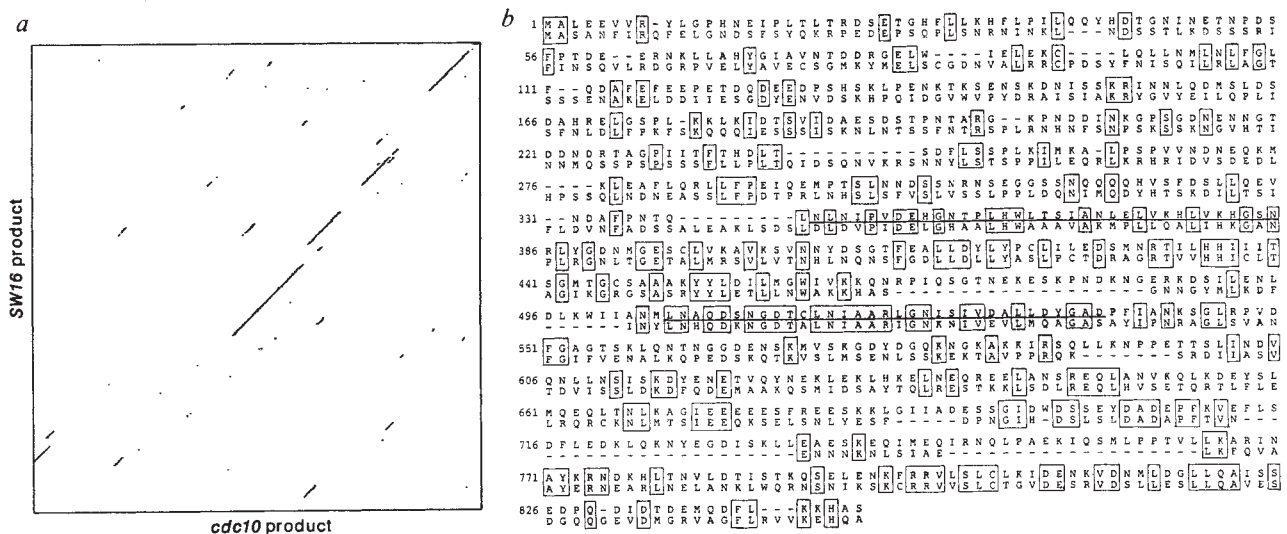


Fig. 2 *a*, DIAGON¹⁸ comparison of the SWI6 protein and the *cdc10* protein. *b*, Alignment of the sequence of the SWI6 product with that of the *cdc10* product using the ALIGN program¹⁹. The clearly significant similarity between the SWI6 and *cdc10* products begins at position 350, and shows 138 identities out of a possible 420 (33%) and has an alignment score of 29. The amino-terminal residues have a 16% match overall and an alignment score of 2. Note the *S. cerevisiae* convention of denoting the wild-type gene in upper case letters and the mutant gene in lower case. Solid lines, the 33-amino-acid repeats also found in *Notch* and *lin-12*.

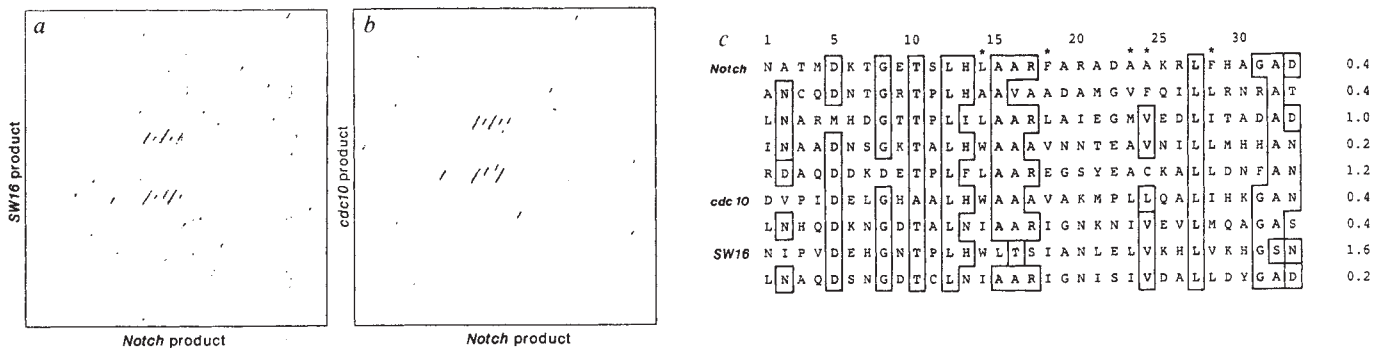


Fig. 3 *a*, *b*, DIAGON plots of SWI6 or *cdc10* versus *Notch*. *c*, Amino-acid sequences that form the 33-amino-acid repeats in the *Notch*, SWI6 and *cdc10* products. Identical residues are enclosed with shaded boxes and conservative substitutions are enclosed with open boxes. *, Positions occupied largely by hydrophobic residues. Scores for each sequence from a TEMPLATE search²⁶, specifying similarity at positions 5, 8, 10, 12, 15, 16, 23, 24, 27, 28, 32 and 33 are listed at the right. The repeats within the *lin-12* sequence (K. M. Weston and I. S. Greenwald, personal communication) have scores in the range of those listed. No other sequences were found that scored <4.0 in a search through 3,700 proteins.

plot that most of the similarity resides in the C-terminal two-thirds of the proteins. An alignment¹⁹ of the two sequences (Fig. 2b) shows that 138 of the C-terminal 420 amino acids of the *cdc10* protein are identical to those of SWI6 when nine gaps are introduced. This 33% similarity is significantly lower than that observed with many functionally homologous genes, and would be very difficult to detect by hybridization. However, this level of similarity has been observed between numerous pairs of genes known to be related²⁰. There is little or no similarity in the N-terminal third of the SWI6 and *cdc10* proteins, suggesting that these regions could have completely unrelated functions. However, subcloning experiments have shown that the part of the *cdc10* gene encoding the N-terminal half of the product is not essential for *cdc10* activity. Thus, the C-terminal half of the gene encodes all the information required for *cdc10* function, and possesses most of the similarity with SWI6. This is consistent with the possibility that SWI6 and *cdc10* have similar functions.

The question of functional homology has been addressed in two ways. First, complementation tests were performed in which

multi-copy plasmids bearing the *cdc10* or SWI6 genes were transformed into *swi6* and *cdc10* mutants respectively and transformants were screened for rescue of the mutant phenotype. No evidence of rescue has been obtained. Second, we tested whether SWI6, like *cdc10*, is an essential gene. Figure 1 shows a deletion mutant in which 80% of the SWI6 coding sequences have been replaced by the *TRP1* gene. This construction was substituted for one copy of SWI6 in a diploid strain, and then the diploid was sporulated. The resulting haploid progeny were all viable, and the two TRP⁺ haploids from each tetrad displayed Swi⁻ phenotypes similar to that of *swi6-399*. Thus, the deletion of SWI6 does not produce a lethal mutation.

This suggests either that SWI6 encodes a nonessential activity, or that there is more than one gene encoding this activity. The duplication of essential genes is not uncommon in *S. cerevisiae*²¹⁻²³, and indeed SWI4 and SWI6 may be such a pair. Both SWI4 and SWI6 are required for CACGA4-specific activation of transcription, and though both single mutants are viable, the *swi4*, *swi6* double mutant is not viable⁹. Furthermore, the

SWI4 gene on a single copy plasmid can complement a *swi6* mutation (see Fig. 1 legend). Thus it would appear that *SWI4* and *SWI6* have an overlapping and essential function in *S. cerevisiae*. This could be the same essential function carried out by *cdc10* in *Schiz. pombe*.

The equally interesting alternative is that *SWI6* and *cdc10* have completely different functions, and instead share some other feature of their structure, metabolism or regulation. Because both proteins are required at the start of the cell cycle, their sequence similarity could reflect the way in which their activities are linked to start. If this is the case there may be a family of start-associated gene products that share this similarity of sequence.

The central one-third of the *SWI6* and *cdc10* proteins is the most conserved segment of the two genes. In this region there is a 33-amino-acid repeat which can also be found tandemly repeated in the cytoplasmic domain of the *Notch* gene of *Drosophila melanogaster*. This conserved sequence was first observed as an eight-residue repeat in the *Notch* protein^{24,25}. It was then shown to be a more extensive repeat with similarity over a 216-amino-acid stretch of *cdc10* (D. Beach and R. Doolittle, personal communication). DIAGON plots (Fig. 3a and b) detect a repeated sequence, present five times in the cytoplasmic domain of *Notch* (positions 1,944 to 2,109) and twice in the central regions of both *SWI6* and *cdc10*. The amino-acid sequences that form the repeated pattern in all three proteins are listed below. A TEMPLATE²⁶ search was done on the basis of these nine examples. Only one other protein was found that contains a set of clearly related repeats: the product of the *lin-12* gene of the nematode *Caenorhabditis elegans*.

Many similarities between the *Notch* and *lin-12* genes have been observed, but it is not clear what these genes may have in common with *SWI6* and *cdc10*. *Notch* and *lin-12* are involved in the determination of cell fates^{27,28}. In addition, the two proteins are structurally similar. Both gene products contain two types of cysteine-rich repeats, followed by a membrane spanning domain, and long cytoplasmic extensions^{24,25,29,30}. It is within the putative cytoplasmic domains that the 33-amino-acid repeat can be found, tandemly repeated and in the same relative position to the membrane spanning region in both proteins.

Internal repeats occur frequently in polypeptide chains³¹. The presence of the same repeat in *SWI6*, *cdc10*, *Notch* and *lin-12* suggests that there may be a common mechanism underlying their activities. The fact that this motif is conserved in a set of evolutionarily distant organisms suggests that it plays an important role in the cell and that it will be found in other organisms.

We thank K. Weston and I. Greenwald for communicating the sequence of *lin-12* before publication; D. Beach and R. Doolittle for pointing out the similarity between *Notch* and *cdc10*; P. Fantes for transforming pLB84 into a *cdc10* mutant of *Schiz. pombe*; and C. Hodgman, A. Lesk, C. Chothia, S. Brenner, K. Weston, R. Staden and D. Bradley for invaluable discussions and advice. L.B. was funded by an NIH postdoctoral fellowship.

17. Draetta, G., Brizuela, L., Potashkin, J. & Beach, D. *Cell* **50**, 319-325 (1987).
18. Staden, R. *Nucleic Acids Res.* **10**, 2951-2961 (1982).
19. Orcutt, B. C., Dayhoff, M. O., George, D. A. & Barker, W. C. *User's Guide for the Alignment Score Program* (National Biomedical Research Foundation, Washington DC, 1984).
20. Doolittle, R. F. *Science* **214**, 149-159 (1981).
21. Tatchell, K., Chaleff, D. T., DeFeo-Jones, D. & Scolnick, E. M. *Nature* **309**, 523-527 (1984).
22. Craig, E. A. & Jacobsen, K. *Cell* **38**, 841-849 (1984).
23. Toda, T., Cameron, S., Sass, P., Zoller, M. & Wigler, M. *Cell* **50**, 277-287 (1987).
24. Wharton, K. A., Johansen, K. M., Xu, T. & Artavanis-Tsakonas, S. *Cell* **43**, 567-581 (1985).
25. Kidd, S., Kelley, M. R. & Young, M. W. *Molec. cell. Biol.* **6**, 3094-3108 (1986).
26. Bashford, D., Chothia, C., Lesk, A. *J. molec. Biol.* **196**, 199-216 (1987).
27. Poulson, D. F. *Proc. natn. Acad. Sci. U.S.A.* **23**, 133 (1937).
28. Greenwald, I. S., Sternberg, P. W. & Horvitz, H. R. *Cell* **34**, 435-444 (1983).
29. Greenwald, I. S. *Cell* **43**, 583-590 (1985).
30. Greenwald, I. S. *Bioessays* **6**, 70-73 (1987).
31. Barker, W. C., Ketcham, L. K. & Dayhoff, M. O. *Atlas of Protein Sequence and Structure*, vol. 5, Suppl. 3, 359-362 (National Biomedical Research Foundation, Washington DC, 1978).
32. MacKay, V. L. *Meth. Enzym.* **101**, 325-343 (1983).
33. Rothstein, R. *Meth. Enzym.* **101**, 202-211 (1983).
34. Southern, E. M. *J. molec. Biol.* **98**, 503-517 (1975).
35. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
36. Henikoff, S. *Gene* **28**, 351-359 (1984).

Inhibition of retroviral protease activity by an aspartyl proteinase inhibitor

Iyoko Katoh*, Teruo Yasunaga†, Yoji Ikawa* & Yoshiyuki Yoshinaka‡

* Laboratory of Molecular Oncology, Tsukuba Life Science Center, Institute of Physical and Chemical Research, 3-1 Koyadai, Yatabe, Ibaraki 305, Japan

† Computer Center, Institute of Physical and Chemical Research, Wako, Saitama 351-01, Japan

‡ Japan Immunoresearch Laboratories Ltd., 17-5 Sakae, Takasaki, Gunma 370, Japan

Retrovirus protease is an enzyme that cleaves *gag* and *gag-pol* precursor polyproteins into the functional proteins of mature virus particles^{1,2}. The correct processing of precursor polyproteins is necessary for the infectivity of virus particles: *in vitro* mutagenesis which introduces deletions into the murine leukaemia virus genome produces a protease-defective virus of immature core form and lacking infectivity^{3,4}. A therapeutic drug effective against disease caused by retrovirus proliferation could likewise interfere with virus maturation. The primary structure has so far been determined for the protease of avian myeloblastosis virus⁵, and of murine⁶, feline⁷ and bovine⁸ leukaemia viruses. Amino acid sequencing of the retrovirus proteases, either after their purification or from prediction from the nucleotide sequence, shows that they possess the Asp-Thr-Gly sequence characteristic of the aspartyl proteinases. In this report we show that retrovirus proteases belong to the aspartyl proteinase group and demonstrate an inhibition by the aspartyl proteinase-specific inhibitor, pepstatin A, on the activity of bovine leukaemia, Moloney murine leukaemia and human T-cell leukaemia virus proteases.

Alignment of the amino acid sequence of retrovirus enzyme protease and the sequence deduced from the cloned viral DNA shows that the protease is encoded by a 5' region of the *pol* gene⁶⁻⁸, though avian viral protease is an exception encoded by a 3' portion of the *gag* gene⁵. It was proposed that synthesis of the proteases could be achieved by suppression of the *gag* termination signal, employing either a readthrough^{6,7} or ribosomal frameshifting⁸ mechanism. Homologous sequences are found in similar locations on other retroviral genomes, including human immunodeficiency virus (HIV) (ref. 9), human T-cell leukaemia virus, types I (refs 10, 11) and II (ref. 12) (HTLV-I and II), and mouse mammary tumour virus (MMTV) (ref. 13), and these viruses probably also possess a protease for generation of mature viral proteins. Among retrovirus proteases, including both those already identified and those predicted from nucleotide sequence analysis, amino acid sequences are strictly con-

Received 16 June; accepted 24 August 1987.

1. Kostricken, R. & Heffron, F. *Cold Spring Harb. Symp. quant. Biol.* **48**, 89-96 (1984).
2. Jensen, R. E. & Herskowitz, I. *Cold Spring Harb. Symp. quant. Biol.* **48**, 97-104 (1983).
3. Jensen, R. E., Sprague, G. F. & Herskowitz, I. *Proc. natn. Acad. Sci. U.S.A.* **80**, 3035-3039 (1983).
4. Siliciano, P. G. & Tatchell, K. *Cell* **37**, 967-978 (1984).
5. Miller, A. M., MacKay, V. L. & Nasmyth, K. A. *Nature* **314**, 598-603 (1985).
6. Nasmyth, K. A. *Nature* **302**, 670-676 (1983).
7. Nasmyth, K. *Cell* **42**, 225-235 (1985).
8. Breeden, L. & Nasmyth, K. *Cold Spring Harb. Symp. quant. Biol.* **50**, 643-650 (1985).
9. Breeden, L. & Nasmyth, K. *Cell* **48**, 389-397 (1987).
10. Bennetzen, J. L. & Hall, B. D. *J. biol. Chem.* **257**, 3026-3031 (1982).
11. Lipman, D. J. & Pearson, W. R. *Science* **227**, 1435-1441 (1985).
12. Aves, S. J., Durkacz, B. W., Carr, A. & Nurse, P. *EMBO J.* **4**, 457-463 (1985).
13. Nurse, P. & Bissett, Y. *Nature* **292**, 558-560 (1981).
14. Beach, D., Durkacz, B. & Nurse, P. *Nature* **300**, 706-709 (1982).
15. Boother, R. & Beach, D. *Molec. cell. Biol.* **6**, 3523-3530 (1986).
16. Lee, M. G. & Nurse, P. *Nature* **327**, 31-35 (1987).

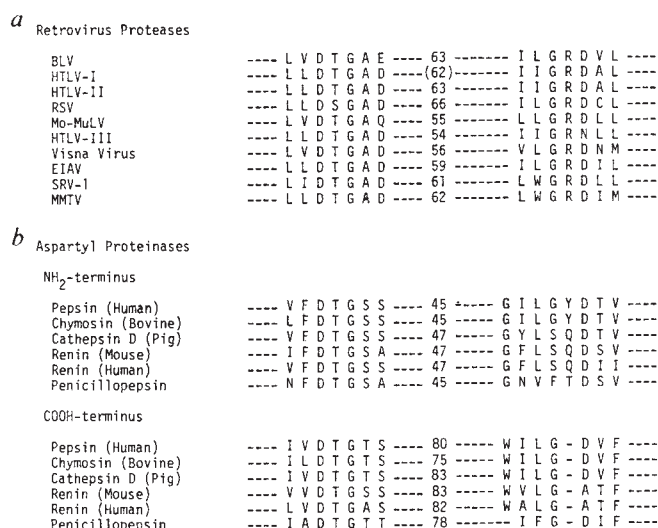


Fig. 1 Alignment of the conserved amino-acid sequences between retrovirus proteases and aspartyl proteinases. **a**, Amino-acid sequences in the two stretches conserved among retrovirus proteases. The sequence comparison between BLV, HTLV-I and II, Rous sarcoma virus (RSV), Moloney murine leukaemia virus (Mo-MuLV), HTLV-III, and Visna virus is described in ref. 14. The sequence data for equine infectious anaemia virus (EIAV), simian retrovirus (SRV) and mouse mammary tumour virus (MMTV) were from refs 29, 30 and 13 respectively. The numbers of the amino-acid residues found between the two stretches are also given. **b**, Sequences found in the NH₂- and COOH-terminal domains of aspartyl proteinases related to the sequences shown in **a**. Amino-acid sequences¹⁶ of pepsin from human, chymosin from bovine, cathepsin D from pig, renin from mouse and human, and penicillopepsin, were aligned. Letter codes used for amino acids are: A, alanine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; I, isoleucine; L, leucine; M, methionine; N, asparagine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; Y, tyrosine.

served in two stretches, represented by -L-V-D-T-G-A- and -I-L-G-R-D- (ref. 14), as shown in Fig. 1a. Conserved sequences are also found in the enzyme group of aspartyl proteinases, formerly referred to as acid proteases or carboxyl proteases (Fig. 1b), which typically have two catalytically important sequences, one at the amino-terminus and the other at the carboxyl-terminus, determined as -F-D-T-G-S- and -I(L or V)-V-D-T-G respectively^{15,16}. Because retrovirus proteases also contain a unique core sequence, -D-T-G-, at the active site, as predicted earlier¹⁷, we compared their catalytic features with the aspartyl proteinases. Pepstatin A (ref. 18), an N-acylpentapeptide (isovaleryl-Val-Val-Sta-Ala-Sta, where Sta represents statine, 4-amino-3-hydroxy-6-methyl-heptanoic acid) isolated from *Streptomyces testaceus*, exerts a strong inhibition on aspartyl proteinases by acting as a substrate analogue. As shown in Fig. 2a, purified bovine leukaemia virus (BLV) protease activity was inhibited by pepstatin A at 0.5 mM in the assay system. On the basis of the reduction of substrate (Pr65^{gag}) band intensity, up to 90% inhibition of protease digestion was observed. Assays of protease activity using whole virus preparations of Moloney MuLV (Fig. 2b) and HTLV-I (Fig. 2c) also showed inhibition by pepstatin A.

To examine whether this inhibition resulted from the binding of pepstatin A with the protease, pepstatin A affinity column chromatography was carried out. When a detergent-disrupted BLV preparation was applied to the column, no protease was detectable in the flow-through fractions (Fig. 3a, lanes 1-6) but was eluted by 1.0M NaCl (lane 7). In the subsequent fractionation with reverse-phase high-performance liquid chromatography (RP-HPLC), the protease was eluted by 52-55%

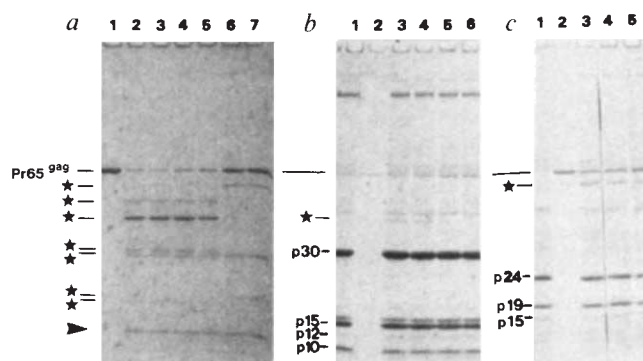


Fig. 2 SDS-PAGE showing the effect of pepstatin A on protease activity of **a**, BLV, **b**, Mo-MuLV, and **c**, HTLV. **a**, BLV protease was purified from the virus produced by Tb1Lu cells derived from bat lung. The protease activity was examined using Gazdar murine sarcoma virus (Gz-MSV) Pr65^{gag} (lane 1) as the substrate⁸. Pepstatin A was contained in the assay reactions at the concentrations of 0 μ M (lane 2), 1 μ M (lane 3), 10 μ M (lane 4), 100 μ M (lane 5), 500 μ M (lane 6) and 1 mM (lane 7). The arrowhead indicates the BLV protease, stained with Coomassie brilliant blue (CBB) R250. Detectable cleavage products from Pr65^{gag} are indicated by asterisks. **b**, Mo-MuLV was isolated from MJD-54 cell⁶ culture. The protease activity was detected by incubating detergent-disrupted whole virus (lane 1) with Pr65^{gag} (lane 2). Pepstatin A was contained in the assay at 0 μ M (lane 3), 100 μ M (lane 4), 500 μ M (lane 5) and 1 mM (lane 6). Mo-MuLV gag proteins p30, p15, p12 and p10 are shown. **c**, HTLV-I was obtained from MT2 cells. Whole virus (lane 1) was incubated with Pr65^{gag} (lane 2) in the presence of pepstatin A at 0 μ M (lane 3), 500 μ M (lane 4) and 1 mM (lane 5). HTLV gag proteins p24, p19 and p15 are indicated. **Methods.** Virus purification procedure was described previously²⁷. BLV protease was prepared to 95% purity by PR-HPLC (ref. 18). BLV protease (0.2 μ g), Mo-MuLV particles (30 μ g) or HTLV-I particles (30 μ g) were incubated with Gz-MSV containing Pr65^{gag} (3 μ g, 0.5 μ g and 2 μ g, respectively) in buffer A (20 mM PIPES, pH 7.0, 0.5% Nonidet P-40 (NP-40) and dithiothreitol). Pepstatin A was dissolved in dimethyl sulphoxide (DMSO) and added to the reactions to give the final concentrations above. Each reaction contained 2% DMSO. There was a slight precipitation of pepstatin A under these conditions. After incubation for 16 h at 24 $^{\circ}$ C, cleavage of Pr65^{gag} was analysed by SDS-PAGE incorporating an 8-18% polyacrylamide gradient.

acetonitrile⁸, and was seen to be the dominant band on SDS polyacrylamide gel electrophoresis (PAGE), purified up to 90% (Fig. 3b, lane 11).

We tested other protease inhibitors, including serine proteinase inhibitors (leupeptin, soybean trypsin inhibitor and chymostatin) and a cysteine proteinase inhibitor (antipain). None of those compounds gave significant inhibition of purified BLV protease at a concentration of 1 mM.

In several aspartyl proteinases, including pepsin¹⁹ and mouse renin²⁰, the aspartyl residue located in the carboxy-terminal active site can be esterified by diazoacetyl norleucine methyl ester (DAN) in the presence of cupric ions, with a concomitant inactivation of the enzyme. Purified BLV protease (1 μ M) was incubated with DAN (80 μ M) in the presence or absence of Cu²⁺ (10 μ M) ions (ref. 20). The reaction was stopped with EDTA (0.1 mM) and assayed for protease activity. After the reaction with DAN alone, the protease was inhibited by about 20% and further inactivated by 50% on addition of Cu²⁺ ions (data not shown). The effect of cupric ions alone was negligible. These results indicate that retrovirus protease might also have a catalytically essential aspartyl residue analogous to the COOH-terminal of aspartyl proteinases.

Despite these similarities, retrovirus proteases differ from aspartyl proteinases in several respects. One of their notable features is a stringent substrate specificity: retrovirus proteases

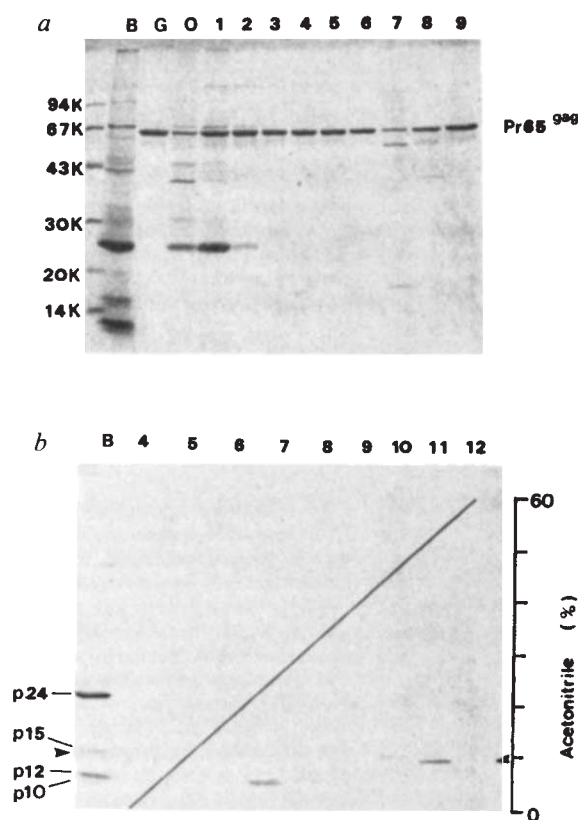


Fig. 3 *a*, SDS-PAGE showing protease in the fractions of pepstatin A affinity column chromatography of BLV solubilized with NP-40. The result of the assays in the starting material (lane 0), the flow-through fractions (lanes 1–6) and the 1.0 M NaCl eluates (lanes 7–9) are shown. Whole BLV proteins and the substrate Gz-MSV Pr65^{gag} are shown in lanes B and G, respectively. Molecular markers are shown on the left. *b*, SDS-PAGE analysis of the proteins adsorbed to pepstatin A-agarose column (*a*, lane 7) and purified by RP-HPLC. Lanes 4–12 are fractions eluted with the acetonitrile gradient. Virus proteins (lane B) p40, p19, p15 and p10 are shown. Arrowheads denote the protease bands on the SDS-PAGE gel.

Methods. BLV (10 mg) was disrupted with 0.5% NP-40, dialysed against buffer A, and clarified by centrifugation at 12,000 r.p.m. for 15 min. The sample was applied on to a pepstatin A-coupled agarose (SIGMA) column (1.5 × 5 cm) equilibrated with 20 mM PIPES buffer with dithiothreitol. After extensive washing, materials adsorbed to the column were eluted by 1.0 M NaCl in PIPES buffer. All procedures were performed at 4 °C. Assay for protease activity was as described in Fig. 2 legend, using Pr65^{gag}. Fraction 7, which contained the protease activity, was further fractionated by RP-HPLC on a micro-Bondapak C₁₈ column (Waters Associates). The gradient used was 0–60% acetonitrile over a period of 40 min at a constant flow of 1.0 ml min⁻¹. 500 µl of each fraction (5 ml) was analysed for protein composition by SDS-PAGE.

from AMV (ref. 21), BLV (ref. 8), FeLV (ref. 5) and HTLV-I cleave mouse *gag* precursor Pr65^{gag} at specific sites only. No other proteins so far examined can act as a substrate for retrovirus proteases unless they are first denatured²¹. Pepsin digests Pr65^{gag} into small fragments of relative molecular masses (*M_r*) below 5,000 (5K).

In the inhibition experiment described in Fig. 2, pepstatin A was less effective against retrovirus proteases than it is generally against aspartyl proteinases. The concentration for fifty percent inhibition (IC₅₀) or pepstatin A for pepsin, cathepsin D and renin is 1.4×10^{-8} M, 8.8×10^{-9} M and 6.6×10^{-6} M respectively²². The IC₅₀ values for retrovirus proteases were in the

range 1.0×10^{-4} M– 5.0×10^{-4} M in the assay system used in this study. This result could reflect the strict substrate specificity of retrovirus proteases.

Optimum pHs for retrovirus proteases range from 5.5–7.2 (refs 8, 23), whereas typical aspartyl proteinases, such as pepsin and chymosin, are active at pH 2.0–4.0 (ref. 24), although renin, an aspartyl protease which specifically cleaves the prehormone angiotensinogen into angiotensin I, has a pH optimum of 5.5–7.0 (ref. 25). To test the affinity of renin for the retrovirus protease substrate, human renin produced by recombinant DNA (a gift from K. Murakami) was incubated with Pr65^{gag}. No cleavage product was detected after prolonged incubation for 24 hours under conditions optimal for digestion of renin substrate.

The aspartyl proteinases, pepsin, chymosin, cathepsin D and renin, consist of some three hundred amino acid residues and have about 30% sequence homology. A bilobal structure that accommodates the two catalytically essential aspartyl residues in a symmetrical arrangement has been proposed²⁶. But the MuLV and BLV proteases are 125 and 126 amino acids long respectively, and other retrovirus proteases are expected to be similar in size. Moreover, sequence homology found between retrovirus proteases and aspartyl proteinases is limited to the two stretches described above. Therefore, retrovirus proteases could possibly act on their substrates by assuming dimeric structures. Indeed, MuLV protease (relative molecular mass 14,000 (14K)), has an apparent molecular weight 20–24K as judged by its activity after elution from a Sephadex G-75 column²⁷. Retrovirus proteases might thus belong to a special category of proteolytic enzymes, with a catalytic mechanism in common with the aspartyl proteinases.

In this study, pepstatin A was found to be an inhibitor of retrovirus proteases. It has been reported that Moloney murine sarcoma virus focus formation was reduced by adding pepstatin to the culture medium²⁸, possibly because virus proliferation was impaired. This observation might be explained by the effect of pepstatin in preventing virus maturation. Development of retrovirus protease inhibitors with strict specificity would provide valuable antiretroviral agents.

We thank J. Clarkin for preparation of this manuscript.

Note added in proof: A three-dimensional structure of retroviral proteases has recently been proposed by Pearl and Taylor (*Nature* 329, 531–534; 1987). Their hypothetical model is consistent with our conclusion drawn by biochemical characterization.

Received 16 June; accepted 10 September 1987.

- Yoshinaka, Y. & Luftig R. B. *Cell* 12, 709–719 (1977).
- Von der Helm, K. *Proc. natn. Acad. Sci. U.S.A.* 74, 911–915 (1977).
- Crawford, S. & Goff, S. P. *J. Virol.* 53, 899–907 (1985).
- Katoh, I. *et al. Virology* 145, 280–292 (1985).
- Sauer, R. T., Allen, D. D. W. & Niall, H. D. *Biochemistry* 20, 3784–3791 (1981).
- Yoshinaka, Y., Katoh, I., Copeland, T. D. & Oroszlan, S. *Proc. natn. Acad. Sci. U.S.A.* 82, 1618–1622 (1985).
- Yoshinaka, Y., Katoh, I., Copeland, T. D. & Oroszlan, S. *J. Virol.* 55, 870–873 (1985).
- Yoshinaka, Y., Katoh, I., Copeland, T. D. & Oroszlan, S. *J. Virol.* 57, 826–832 (1986).
- Ratner, L. *et al. Nature* 313, 227–284 (1986).
- Inouye, J. *et al. Virology* 150, 187–195 (1986).
- Hiramatsu, K., Nishida, J., Naito, A. & Yoshikura, H. *J. gen. Virol.* 68, 213–218 (1987).
- Shimotohno, K. *et al. Proc. natn. Acad. Sci. U.S.A.* 82, 3101–3105 (1985).
- Moore, R., Dixon, M., Smith, R., Peters, G. & Dickson, C. *J. Virol.* 61, 480–490 (1987).
- Yasunaga, T., Sagata, N. & Ikawa, Y. *FEBS Lett.* 199, 145–150 (1986).
- Miyazaki, H. *et al. Proc. natn. Acad. Sci. U.S.A.* 81, 5999–6003 (1984).
- Misono, K., Chang, J. J. & Inagami, T. *Proc. natn. Acad. Sci. U.S.A.* 79, 4858–4862 (1982).
- Toh, H., Ono, M., Saigo, K. & Miyata, T. *Nature* 315, 691 (1985).
- Umezawa, H. *et al. J. Antibiot., Tokyo* 23, 259–262 (1970).
- Ragagopalan, T. G., Stein, W. H. & Moore, S. *J. biol. Chem.* 241, 4295–4297 (1966).
- Misono, K. S. & Inagami, T. *Biochemistry* 19, 2616–2622 (1980).
- Yoshinaka, Y. & Luftig, R. B. *Virology* 111, 239–250 (1981).
- Aoyagi, T. *et al. J. Antibiot., Tokyo* 26, 539–541 (1973).
- Yoshinaka, Y. & Luftig, R. B. *Biochem. biophys. Res. Commun.* 76, 54–56 (1977).
- Foltmann, B. & Pederson, V. B. in *Acid Proteases: Structure, Function and Biology* (ed. Tang, J.) 3–22 (Plenum, New York, 1977).
- Yokosawa, H., Holladay, L. A., Inagami, T., Haas, E. & Murakami, K. *J. Biol. Chem.* 225, 2498–2502 (1980).
- James, M. N. G. & Sielecki, A. R. *J. molec. Biol.* 163, 299–361 (1983).
- Yoshinaka, Y. & Luftig, R. B. *J. gen. Virol.* 48, 329–340 (1980).
- Yuasa, Y., Shimajo, H., Aoyagi, T. & Umezawa, H. *J. natn. Cancer Inst.* 54, 1255–1256 (1975).
- Stephens, R. M., Casey, J. W. & Rice, N. R. *Science* 231, 589–594 (1986).
- Power, M. D. *et al. Science* 231, 1567–1572 (1986).